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

First, we need to find the likelihood:

$$p(y_1,y_2,\ldots,y_n|\mu,\omega)=\frac{1}{(2\pi)^{\frac{n}{2}}}\omega^{\frac{n}{2}}exp\left(-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right)=\frac{1}{(2\pi)^{\frac{n}{2}}}\omega^{\frac{n}{2}}exp\left(-\frac{\omega}{2}\left[n(\mu-\bar{y})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2\right]\right)$$

From the question we have the prior on μ and the prior on ω, Conjugate prior on μ,ω is as following:

$$\begin{aligned}p(\mu,\omega|\mu_0,k_0,\alpha_0,\beta_0) &= N\left(\mu_0,\frac{1}{k_0\omega}\right)*Gamma(\omega|\alpha_0,\beta_0)\\ &=\frac{1}{\frac{\Gamma(\alpha_0)}{\beta_0{}^{\alpha_0}}\left(\frac{2\pi}{k_0}\right)^{\frac{1}{2}}}\omega^{\frac{1}{2}}exp\left(-\frac{k_0\omega}{2}(\mu-\mu_0)^2\right)\omega^{\alpha_0-1}exp(-\omega\beta_0)\\ &=\frac{1}{\frac{\Gamma(\alpha_0)}{\beta_0{}^{\alpha_0}}\left(\frac{2\pi}{k_0}\right)^{\frac{1}{2}}}\omega^{\alpha_0-\frac{1}{2}}exp\left(-\frac{\omega}{2}\left[k_0(\mu-\mu_0)^2+2\beta_0\right]\right)\end{aligned}$$

Posterior distribution is proportional to the product of likelihood and conjugate prior:

$\mathrm{p}(\mu, \omega | y_1, y_2, \dots, y_n) \propto \mathrm{p}(\mu, \omega | \mu_0, k_0, \alpha_0, \beta_0) * \mathrm{p}(y_1, y_2, \dots, y_n | \mu, \omega) \propto \frac{1}{\frac{\Gamma(\alpha_0)}{\beta_0^{\alpha_0}}\left(\frac{2\pi}{k_0}\right)^{\frac{1}{2}}} \omega^{\alpha_0 - \frac{1}{2}} exp\left(-\frac{\omega}{2}[k_0(\mu - \mu_0)^2 + 2\beta_0]\right) \frac{1}{(2\pi)^{\frac{n}{2}}} \omega^{\frac{n}{2}} exp\left(-\frac{\omega}{2}[n(\mu - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2]\right)$

$$\propto \omega^{\frac{1}{2}} \omega^{\frac{n}{2}+\alpha_0-1} exp\left(-\frac{\omega}{2}[k_0(\mu - \mu_0)^2 + 2\beta_0] - \frac{\omega}{2}\left[n(\mu - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2\right]\right)$$

$$\propto \omega^{\frac{1}{2}} \omega^{\frac{n}{2}+\alpha_0-1} exp\left(-\frac{\omega}{2}\left[k_0(\mu - \mu_0)^2 + n(\mu - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\beta_0\right]\right)$$

$**k_0(\mu - \mu_0)^2 + n(\mu - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2 = (k_0 + n)(\mu - \mu_n)^2 + \frac{k_0 n(\bar{y} - \mu_0)^2}{k_0 + n} + \sum_{i=1}^{n}(y_i - \bar{y})^2$

where $\mu_n = \frac{k_0\mu_0 + n\bar{y}}{k_0 + n}$

So,

$$\begin{aligned} \mathrm{p}(\mu, \omega | y_1, y_2, \dots, y_n) \\ &\propto \omega^{\frac{1}{2}} exp\left(-\frac{\omega}{2}[(k_0 + n)(\mu - \mu_n)^2]\right) \\ &* \omega^{\frac{n}{2}+\alpha_0-1} exp(-\beta_0\omega) exp\left(-\frac{\omega}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right) exp\left(-\frac{\omega}{2}\frac{k_0 n(\bar{y} - \mu_0)^2}{k_0 + n}\right) \\ &\propto \mathrm{N}(\mu | \mu_n, ((k_n)\omega)^{-1}) * Gamma(\omega | \alpha_n, \beta_n) \end{aligned}$$

Normal Gamma distribution, where

$$\mu_n = \frac{k_0\mu_0 + n\bar{y}}{k_0 + n}$$

$$k_n = k_0 + n$$

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

$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{k_0 n(\bar{y} - \mu_0)^2}{2(k_0 + n)}$$

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

Solution

$\mathrm{p}(\mu|\omega, y_1, y_2, \dots, y_n) = p(y_1, y_2, \dots, y_n|\mu, \omega)\, p(\mu|\omega)$

First, we need to find $p(y_1, y_2, \dots, y_n|\mu, \omega)$ which is equal to

$$p(y_1, y_2, \dots, y_n|\mu,\omega) = \frac{1}{(2\pi)^{\frac{n}{2}}}\omega^{\frac{n}{2}} exp\left(-\frac{\omega}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) = \frac{1}{(2\pi)^{\frac{n}{2}}}\omega^{\frac{n}{2}} exp\left(-\frac{\omega}{2}\left[n(\mu-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2\right]\right)$$

Next, we need to find $p(\mu|\omega)$ which is equal to:

$$p(\mu|\omega) = \frac{(\omega k_0)^{\frac{1}{2}}}{\sqrt{2\pi}} \, exp\left(-\frac{k_0\omega}{2}(\mu-\mu_0)^2\right)$$

So,

$$\begin{aligned}
p(\mu|\omega, y_1, y_2, \dots, y_n) &= \frac{1}{(2\pi)^{\frac{n}{2}}}\omega^{\frac{n}{2}} exp\left(-\frac{\omega}{2}\left[n(\mu-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2\right]\right)\frac{(\omega k_0)^{\frac{1}{2}}}{\sqrt{2\pi}} \, exp\left(-\frac{k_0\omega}{2}(\mu-\mu_0)^2\right) \\
&\propto \exp\left(-\frac{\omega}{2}[n(\mu-\bar{y})^2 - k_0(\mu-\mu_0)^2]\right) \\
&\propto \exp\left(-\frac{\omega}{2}[n(\mu^2-2\mu\bar{y}+\bar{y}^2) + k_0(\mu^2-2\mu\mu_0+\mu_0{}^2)]\right) \\
&\propto \exp\left(-\frac{\omega}{2}[(n+k_0)\mu^2 - 2\mu(n\bar{y}+k_0\mu_0)]\right) \propto exp\left(\frac{-\omega(n+k_0)(\mu-\frac{n\bar{y}+k_0y_0}{n+k_0})}{2}\right)
\end{aligned}$$

The above function is the kernel of normal distribution. Therefore, conditional posterior distribution of μ is:

$$p(\mu|\omega, y_1, y_2, \dots, y_n) \sim Normal\left(\frac{n\bar{y}+k_0y_0}{n+k_0}, \frac{1}{\omega(n+k_0)}\right)$$

Reasonable interpretation on parameters $\mu_0, k_0, \alpha_0, \beta_0$

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

Marginal distribution over μ is obtained through following:

$$p(\mu) = \int p(\mu,\omega)d\omega = \int \frac{1}{\frac{\Gamma(\alpha_0)}{\beta_0{}^{\alpha_0}}\left(\frac{2\pi}{k_0}\right)^{\frac{1}{2}}}\omega^{\alpha_0+\frac{1}{2}-1} exp\left(-\frac{\omega}{2}[k_0(\mu-\mu_0)^2 + 2\beta_0]\right) d\omega \propto \int \omega^{\alpha_0+\frac{1}{2}-1} exp\left(-\omega\left[\frac{k_0}{2}(\mu-\mu_0)^2 + \beta_0\right]\right) d\omega$$

Under the integral is the kernel of gamma (α, β) distribution where $\alpha = \alpha_0 + \frac{1}{2}$, $\beta = \frac{k_0}{2}(\mu - \mu_0)^2 + \beta_0$.

So,

$$p(\mu) \propto \frac{\Gamma(\alpha)}{\beta^{\alpha}} \propto \beta^{-\alpha} = (\frac{k_0}{2}(\mu-\mu_0)^2 + \beta_0)^{-\frac{(2\alpha_0+1)}{2}} = (1 + \frac{1}{2\alpha_0}\frac{\alpha_0 k_0(\mu-\mu_0)^2}{\beta_0})^{-\frac{(2\alpha_0+1)}{2}}$$

a scaled t distribution where m $= \mu_0$, v $= 2\alpha_0, s^2 = \frac{\beta_0}{\alpha_0 k_0}$

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

$$p(\mu) = \int p(\mu,\omega)d\omega = \int \omega^{\frac{1}{2}} exp\left(-\frac{\omega}{2}[k_n(\mu-\mu_n)^2]\right) * \omega^{\alpha_n - 1} exp\left(-\beta_n \omega\right) d\omega$$
$$= \int \omega^{\alpha_n + \frac{1}{2} - 1} exp(-\omega[\beta_n + \frac{k_n(\mu-\mu_n)^2}{2}])d\omega$$

Under the integral is the kernel of gamma (α, β) distribution

where $\alpha = \alpha_n + \frac{1}{2}$, $\beta = \frac{k_n}{2}(\mu - \mu_n)^2 + \beta_n$.

So,

$$p(\mu) \propto \frac{\Gamma(\alpha)}{\beta^{\alpha}} \propto \beta^{-\alpha} = \left(\frac{k_n}{2}(\mu-\mu_n)^2 + \beta_n\right)^{-\frac{(2\alpha_n+1)}{2}} = \left(1 + \frac{1}{2\alpha_n}\frac{\alpha_n k_n(\mu-\mu_n)^2}{\beta_n}\right)^{-\frac{(2\alpha_n+1)}{2}}$$

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

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

Solution

It is better to first solve the exercise 2.6 and then solve this exercise.

$$p\ (y_{n+1}, y_{n+2}, \ldots, y_{n+m}|y_1, y_2, \ldots, y_n) = \frac{p(Y_{new}\ \&\ Y_{old})}{p(Y_{old})}$$

from the exercise 2.6 we found the marginal distribution over Y. So, we can write

$$\frac{p(Y_{new}\ \&\ Y_{old})}{p(Y_{old})} = \frac{\frac{Z_{n+m}}{Z_0\ (2\pi)^{\frac{n+m}{2}}}}{\frac{Z_n}{Z_0\ (2\pi)^{\frac{n}{2}}}} = \frac{Z_{n+m}}{Z_n(2\pi)^{\frac{m}{2}}} == \frac{\Gamma(\alpha_{n+m})}{\Gamma(\alpha_n)} \frac{{\beta_n}^{\alpha_n}}{{\beta_{n+m}}^{\alpha_{n+m}}} (\frac{k_n}{k_{n+m}})^{\frac{1}{2}} (2\pi)^{\frac{-m}{2}}$$

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

Solution:

$$p(\mu, \omega|Y) = \frac{p(Y, \mu, \omega)}{p(Y)} \rightarrow p(Y) = \frac{p(Y, \mu, \omega)}{p(\mu, \omega|Y)}$$

$$p(Y, \mu, \omega) = NG\left(\left(\mu, \omega \middle| \mu_0, k_0, \alpha_0, \beta_0\right) \prod_{i=1}^{n} N(y_i|\mu, \omega)\right.$$

$$= \frac{1}{Z_0} \omega^{\alpha_0 - \frac{1}{2}} exp\left(-\frac{\omega}{2}[k_0(\mu - \mu_0)^2 + 2\beta_0]\right) * \frac{1}{(2\pi)^{\frac{n}{2}}} \omega^{\frac{n}{2}} exp\left(-\frac{\omega}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

$$= \frac{\omega^{\alpha_n - \frac{1}{2}}}{Z_0\,(2\pi)^{\frac{n}{2}}}\, exp\,\{-\frac{\omega}{2} k_0(\mu - \mu_n)^2 + 2\beta_n\}$$

where $Z_0 = \frac{\Gamma(\alpha_0)}{\beta_0{}^{\alpha_0}} \left(\frac{2\pi}{k_0}\right)^{\frac{1}{2}}$

If we divide the above equation by the posterior distribution:

$$p(y) = \frac{\frac{\omega^{\alpha_n - \frac{1}{2}}}{Z_0\,(2\pi)^{\frac{n}{2}}}\, exp\,\{-\frac{\omega}{2}(k_0(\mu - \mu_n)^2 + 2\beta_n)\}}{\frac{1}{Z_n}\, exp\{\frac{-\omega}{2}(k_0(\mu - \mu_n)^2 + 2\beta_n)\}} = \frac{Z_n}{Z_0\,(2\pi)^{\frac{n}{2}}} = \frac{\Gamma(\alpha_n)}{\Gamma(\alpha_0)} \frac{\beta_0{}^{\alpha_0}}{\beta_n{}^{\alpha_n}} (\frac{k_0}{k_n})^{\frac{1}{2}} (2\pi)^{\frac{-n}{2}}$$

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

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

Solution:

Probability density function of the inverse Wishart is as follows:

$$\frac{|\Lambda|^{\frac{\nu_0}{2}}}{2^{(\nu_0 * d)/2}\Gamma_d(\nu_0/2)} |\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2}tr(\Lambda\Sigma^{-1})}$$

Where Λ and Σ are d*d positive definite matrices, and $\Gamma_d(.)$ is the multivariate gamma function.

With d=1, $\alpha = \frac{\nu_0}{2}$, $\beta = \frac{\Lambda}{2}$, the probability density function of inverse Wishart becomes:

$$\frac{\left|\frac{\Lambda}{2}\right|^{\frac{\nu_0}{2}}}{\Gamma_1(\nu_0/2)} |\Sigma|^{-\frac{\nu_0}{2}-1} e^{-\frac{1}{2}(\Lambda\Sigma^{-1})} = \frac{\beta^{\alpha}}{\Gamma_1(\alpha)}|\Sigma|^{-\alpha-1}e^{-\frac{\beta}{\Sigma}} = \text{Inverse Gamma } (\Sigma;\ \alpha, \beta)$$

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

Posterior distribution can be written as following:

$$p(\mu, \Sigma|y_1, y_2, \ldots, y_n) \propto p(\mu, \Sigma|\mu_0, k_0, \alpha_0, \beta_0) * p(y_1, y_2, \ldots, y_n|\mu, \Sigma)$$

From the question we have:

$$p(\mu, \Sigma|\mu_0, k_0, \alpha_0, \beta_0) \propto |\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2}tr(\Lambda_0\Sigma^{-1})-\frac{k_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)}$$

$$p(y_1, y_2, \ldots, y_n|\mu, \Sigma) = |\Sigma|^{-\frac{n}{2}} exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu))$$

So,

$$p(\mu, \Sigma|y_1, y_2, \ldots, y_n) =$$

$$|\Sigma|^{-\frac{v_0+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0\Sigma^{-1}) - \frac{k_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)) * |\Sigma|^{-\frac{n}{2}} exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu))$$

$$= |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0\Sigma^{-1}) - tr(\Sigma^{-1}\frac{k_0}{2}(\mu-\mu_0)^T(\mu-\mu_0))$$

$$-\frac{1}{2}tr(\Sigma^{-1}\sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu))$$

$$= |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0 + k_0(\mu-\mu_0)^T(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu)\Sigma^{-1})$$

$$= |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0 + k_0(\mu-\mu_0)^T(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\bar{y})^T(y_i-\bar{y})$$

$$+ n(\mu-\bar{y})^T(\mu-\bar{y})\Sigma^{-1})$$

$$= |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0 + S + (k_0+n)\mu^2 - 2\mu(k_0\mu_0 + n\bar{y}) + k_0{\mu_0}^2$$

$$+ n\bar{y}^2)\Sigma^{-1})$$

$$\propto |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0 + S + (\mu - \frac{k_0\mu_0+n\bar{y}}{k_0+n})^T(\mu - \frac{k_0\mu_0+n\bar{y}}{k_0+n})$$

$$+ \frac{k_0 n}{k_0+n}(\bar{y}-\mu_0)^T(\bar{y}-\mu_0))\Sigma^{-1})$$

$$= |\Sigma|^{-\frac{v_0+n+d+1}{2}} exp((-\frac{1}{2}tr(\Lambda_0 + S + \frac{k_0 n}{k_0+n}(\bar{y}-\mu_0)^T(\bar{y}-\mu_0))\Sigma^{-1})$$

$$+ (\mu - \frac{k_0\mu_0+n\bar{y}}{k_0+n})^T\Sigma^{-1}(\mu - \frac{k_0\mu_0+n\bar{y}}{k_0+n})$$

** $k_0(\mu-\mu_0)^T(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\mu)^T(y_i-\mu)) = k_0(\mu-\mu_0)^T(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\bar{y})^T(y_i-\bar{y}) + n(\mu-\bar{y})^T(\mu-\bar{y}))$

** $S = \sum_{i=1}^{n}(y_i-\bar{y})^T(y_i-\bar{y})$

So, the posterior is normal inverse Wishart with the following updated parameters:

$$v_n = v_0 + n$$

$$\Lambda_n = \Lambda_0 + S + \frac{k_0 n}{k_0+n}(\bar{y}-\mu_0)^T(\bar{y}-\mu_0)$$

$$k_n = k_0 + n$$

$$\mu_n = \frac{k_0\mu_0+n\bar{y}}{k_0+n}$$

Interpretations for the prior parameter:

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

From the question we have:

$y|\beta, X \sim$ Normal $(X\beta, (\omega\Lambda)^{-1})$

$\beta \sim$Normal $(\mu, (\omega K)^{-1})$

$\omega \sim$ Gamma (a, b)

We want to find $p(\beta|\omega, y_1, y_2, \ldots, y_n)$

So, we can write

$$\begin{aligned}
p(\beta|\omega, y_1, y_2, \ldots, y_n) &= p(y_1, y_2, \ldots, y_n|\beta, \omega) * p(\beta|\omega) = N(X\beta, (\omega\Lambda)^{-1}) * N(\mu, (\omega K)^{-1}) \\
&\propto \exp(\frac{-(\omega\Lambda)(Y - X\beta)^T(Y - X\beta)}{2})\, exp(\frac{-(\omega K)(\beta - \mu)^T(\beta - \mu)}{2}) \\
&\propto \exp(\frac{-(\omega)[(\Lambda(Y - X\beta)^T(Y - X\beta)) + (K(\beta - \mu)^T(\beta - \mu))]}{2}) \\
&\propto \exp(\frac{-(\omega)[((Y)^T\Lambda Y - 2(Y)^T\Lambda X\beta + (X\beta)^T X\beta \ + \ \beta^T K\beta \ - \ 2\,]}{2}) \\
&\propto exp\,(\frac{-(\omega(K + X^T\Lambda X))\,(\beta \ - \ \frac{X^T\Lambda Y + k\mu}{K + X^T\Lambda X})}{2}
\end{aligned}$$

So, this distribution is normal with mean of $\frac{X^T\Lambda Y + k\mu}{K + X^T\Lambda X}$ and variance of $(\omega(K + X^T\Lambda X))^{-1}$.

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

Solution:

$$p(\omega|Y) \ = \int p(\beta, \omega|y)\, d\beta$$

$$** \ p(\beta, \omega|Y) = p(\beta, \omega) * p(Y|\beta, \omega) = p(\beta|\omega)p(\omega)p(Y|\beta, \omega)$$

p(ω|Y)

$$= \int p(\beta|\omega)p(\omega)p(Y|\beta,\omega)d\beta$$

$$= p(\omega)\int p(\beta|\omega)\, p(Y|\beta,\omega)d\beta \;\; = \omega^{a-1}\exp(-\mathrm{b}\omega)\; \int \left(\frac{\omega \mathrm{K}}{2\pi}\right)^{\frac{1}{2}} exp\left(\frac{-(\omega)(\beta-\mu)^T K(\beta-\mu)}{2}\right)$$

$$* \left(\frac{\omega\Lambda}{2\pi}\right)^{\frac{n}{2}} \exp\left(\frac{-(\omega)(Y-X\beta)^T\Lambda(Y-X\beta)}{2}\right) d\beta$$

$$= \omega^{a+\frac{1}{2}+\frac{n}{2}-1}\exp(-\mathrm{b}\omega)\; \int exp\left(\frac{-(\omega)[(\beta-\mu)^T K(\beta-\mu)+(Y-X\beta)^T\Lambda(Y-X\beta)]}{2}\right) d\beta$$

$$= \omega^{a+\frac{1}{2}+\frac{n}{2}-1}\exp(-\mathrm{b}\omega)\;\; exp\,(\frac{-(\omega)[\mu^T K\mu + Y^T\Lambda \mathrm{Y}]}{2})\; \int exp(\frac{-(\omega)[\beta^T K\beta - 2\mu^T K\beta - 2Y^T\Lambda \mathrm{X}\beta + (\mathrm{X}\beta)^T\Lambda \mathrm{X}\beta]}{2})\, d\beta$$

$$= \omega^{a+\frac{1}{2}+\frac{n}{2}-1}\; exp\,(-\omega[\mathrm{b} + \frac{\mu^T K\mu \; + \; Y^T\Lambda \mathrm{Y}}{2}] \int exp\, \frac{-(\omega)(k+X^T\Lambda X)[(\beta \; - \; \frac{\mu k \; + Y^T\Lambda X}{K+X^T\Lambda X})^2 \; - \; (\frac{\mu k \; + Y^T\Lambda X}{K+X^T\Lambda X})^2]}{2} d\beta$$

$$** \; \mu_n \;\; = \; \frac{\mu k + Y^T\Lambda X}{K + X^T\Lambda X}$$

$$\omega^{a+\frac{1}{2}+\frac{n}{2}-1} exp\,(-(\omega)[b + \frac{1}{2}(\mu^T K\mu \; + \; Y^T\Lambda \mathrm{Y} \; - \; \mu_n{}^T(k+X^T\Lambda X)\mu_n)])$$

A gamma distribution with the updated parameters of

$$a_n \; = \; a + \tfrac{n+1}{2}, b_n \; = \; b + \tfrac{1}{2}(\mu^T K\mu \; + \; Y^T\Lambda \mathrm{Y} \; - \; \mu_n{}^T(k+X^T\Lambda X)\mu_n)$$

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

Solution:

$$\mathrm{p}(\beta|\mathrm{Y}) \; = \; \int p(\beta|\omega, y) * p(\omega|y)d\omega$$

$$= \int (\frac{\omega(K+X^T\Lambda \mathrm{X})}{2\pi})^{\frac{1}{2}}\; exp(\frac{-\;(\beta \; - \; \mu_n)^T\omega(K+X^T\Lambda \mathrm{X})(\beta \; - \; \mu_n)}{2})\; \frac{(b_n)^{a_n}}{\Gamma(\alpha_n)}\,\omega^{\;a_n-1} exp(-\omega\;(b_n))\; \mathrm{d}\omega$$

$$= \;(\frac{(K+X^T\Lambda \mathrm{X})}{2\pi})^{\frac{1}{2}}\; \frac{(b_n)^{a_n}}{\Gamma(\alpha_n)} \int \omega^{\;a_n+\frac{1}{2}-1} exp(-\omega[\,b_n + \frac{1}{2}\;((\beta \; - \; \mu_n)^T(K+X^T\Lambda \mathrm{X})(\beta \; - \; \mu_n))]\; d\omega$$

Kernel of Gamma Distribution

$$= \;(\tfrac{(K+X^T\Lambda \mathrm{X})}{2\pi})^{\frac{1}{2}}\; \tfrac{(b_n)^{a_n}}{\Gamma(\alpha_n)}\; \frac{\Gamma(a_n+\frac{1}{2})}{[b_n+\frac{1}{2}\,((\beta-\mu_n)^T(K+X^T\Lambda \mathrm{X})(\beta-\mu_n))]^{a_n+\frac{1}{2}}} \;\; = (\tfrac{(K+X^T\Lambda \mathrm{X})}{2\pi})^{\frac{1}{2}}\; \tfrac{(b_n)^{a_n}\Gamma(a_n+\frac{1}{2})}{\Gamma(\alpha_n)}\;[b_n +$$

$$\tfrac{1}{2}\;((\beta \; - \; \mu_n)^T(K+X^T\Lambda \mathrm{X})(\beta \; - \; \mu_n))]^{\;-(\frac{2a_n+1}{2})}$$ t distribution with $2a_n$ degrees of freedom.

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

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

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

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

Solution:

$$\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n) \quad , \lambda \sim Gamma(\tau, \tau)$$

$$\mathrm{p}(\Lambda\,|\omega, \beta, y) \propto \mathrm{p}(y|\omega, \beta, \Lambda)\, \mathrm{p}\,(\omega, \beta, \Lambda) \propto \mathrm{p}(y|\omega, \beta, \Lambda)\, \mathrm{p}(\Lambda)\, \mathrm{p}(\omega, \beta)$$

* Λ is independent of ω and β.

$$\propto \lambda_i^{\frac{1}{2}} exp\,(\frac{-\omega\lambda_i(y_i - x_i\beta)}{2})\, \lambda_i^{\tau-1} exp(-\tau\lambda_i) \propto \lambda_i^{\tau+\frac{1}{2}-1} exp((\frac{-\omega(y_i - x_i\beta)}{2} - \tau)\lambda_i)$$

So, Gamma($\tau + \frac{1}{2}, \frac{-\omega(y_i - x_i\beta)}{2} - \tau$).

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

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