Exercise 2.1 Bayesian Inference in a simple Gaussian model

$y_i|\mu,\omega \sim \mathrm{N}(\mu, 1/\omega)$
$\omega \sim \mathrm{Gamma}(\alpha_0, \beta_0)$
$\mu|\omega \sim \mathrm{Normal}(\mu_0, 1/\omega\kappa_0)$

Exercise 2.1:

$p(\mu,\omega|y_1,...,y_n) \propto p(y_1,..,y_n|\mu,\omega)p(\mu|\omega)p(\omega)$
$p(\mu,\omega) = p(\mu|\omega)p(\omega)$
$= \sqrt{\frac{\omega\kappa_0}{2\pi}}\exp\Big\{-\frac{\omega\kappa_0}{2}(\mu-\mu_0)^2\Big\}\frac{\beta_0^{\alpha_0}\omega^{\alpha_0-1}\exp\{-\beta_0\omega\}}{\Gamma\alpha_0}$

$p(\mu,\omega|y_1,...,y_n) \propto \omega^{\alpha_0+\frac{(n+1)}{2}-1}\exp\Big\{-\omega(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2+0.5\sum(y_i-\mu)^2)\Big\}$

It seems that the distribution of $p(\mu,\omega|y_1,...,y_n)$ belongs to the same family of $p(\mu,\omega)$ which is normal-inverse gamma distribution. The updated parameters are:
$\alpha_n = \alpha_0 + \frac{n}{2}$
$\beta_n = \beta_0 + \frac{\kappa_0}{2}(\mu-\mu_0)^2 + 0.5\sum(y_i-\mu)^2$
$\mu_n = \frac{\kappa_0\mu_0+\sum y_i}{\kappa_0+n}$
$\kappa_n = \kappa_0 + n$

Exercise 2.2:

$p(\mu|\omega,y_1,...,y_n) \propto p(y_1,..,y_n|\mu,\omega)p(\mu|\omega)$
$= (\sqrt{\frac{\omega}{2\pi}})^n e^{-\frac{\omega}{2}\sum(y_i-\mu)^2}\sqrt{\frac{\omega\kappa_0}{2\pi}}e^{-\frac{\omega\kappa_0}{2}(\mu-\mu_0)^2}$
$\propto \omega^{\frac{n}{2}+\frac{1}{2}}e^{-\frac{\omega}{2}[\sum(y_i-\mu)^2+\kappa_0(\mu-\mu_0)^2]}$

$p(\omega|y_1,..,y_n) = \int p(\mu,\omega|y_1,...,y_n)d\mu$
$= \int \omega^{\alpha_0+\frac{n}{2}-\frac{1}{2}}e^{-(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2+0.5\sum(y_i-\mu)^2)\omega}d\mu$
$\propto \omega^{\alpha_0+\frac{n}{2}-\frac{1}{2}}\omega^{-\frac{1}{2}}(\kappa_0+n)^{-\frac{1}{2}}e^{-(\beta_0+\frac{1}{2}(\kappa_0\mu_0^2+\sum y_i^2)+\frac{(\kappa_0\mu_0+\sum y_i)^2}{2(\kappa_0+n)})\omega}$

$\omega|y_1,..,y_n \sim \mathrm{Gamma}(\alpha^*,\beta^*)$
$\alpha^* = \alpha_0 + \frac{n}{2}$
$\beta^* = \beta_0 + \frac{1}{2}(\kappa_0\mu_0^2 + \sum y_i^2) + \frac{(\kappa_0\mu_0+\sum y_i)^2}{2(\kappa_0+n)}$

Exercise 2.3:

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

$m = \mu_0$
$\nu = 2\alpha_0$
$s^2 = \frac{\beta_0}{\kappa_0\alpha_0}$

Exercise 2.5:

Posterior predictive distribution:
$p(y_{n+1}, ..., y_{n+m}|y_1, ..., y_m) = \int\int p(y_{n+1}, ..., y_{n+m}|\mu, \omega, y_1, ..., y_m)p(\mu, \omega|y_1, ..., y_m)d\mu d\omega$
$\propto \int\int \omega^{\frac{m}{2}+\alpha_0+\frac{n}{2}-\frac{1}{2}} e^{-\omega(\beta_0+\frac{\kappa_0}{2}(\mu-\mu_0)^2+\frac{1}{2}\sum_1^{n+m}(y_i-\mu)^2)} d\omega d\mu$

Exercise 2.7 Bayesian Inference in a multivariate Gaussian model

Let $\Lambda^{-1} = \Psi, \quad \nu > d-1, \quad \Sigma : d \times d$ p.d. matrix.
$p(\Sigma|\nu, \Psi) = \frac{|\Psi|^{\frac{\nu}{2}}}{2^{\frac{\nu d}{2}}\Gamma_d(\frac{\nu}{2})} |\Sigma|^{-\frac{\nu+d+1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\Psi\Sigma^{-1})}$

When $d = 1$ then the above expression boils down to univariate distribution. Let $\frac{\nu}{2} = \alpha$ and $\frac{\Psi}{2} = \beta$ and $\Sigma = \omega$.

$p(\omega|\alpha, \beta) = \frac{(2\beta)^\alpha}{2^\alpha\Gamma(\alpha)} \omega^{-\alpha-1} e^{-\frac{\beta}{\omega}}$
$= \frac{\beta^\alpha}{\Gamma(\alpha)} \omega^{-\alpha-1} e^{-\frac{\beta}{\omega}}$

The above expression clearly shows that the distribution resembles an inverse gamma distribution with parameters α and β.

Exercise 2.8:

$\Sigma \sim \text{Inv-Wishart}(\nu_0, \Lambda_0^{-1}), \quad \mu|\Sigma \sim N(\mu_0, \frac{\Sigma}{\kappa_0})$
$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{\nu_0+d+2}{2}} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)}$
$y_i \sim N(\mu, \Sigma)$
$p(\underset{\sim}{y}|\mu\Sigma) \propto |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum(\underset{\sim}{y}-\mu)^T\Sigma^{-1}(\underset{\sim}{y}-\mu)}$

$p(\mu, \Sigma|y_1, .., y_n) \propto p(y_1, .., y_n|\mu, \Sigma)p(\mu, \Sigma)$
$\propto |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum(\underset{\sim}{y}-\mu)^T\Sigma^{-1}(\underset{\sim}{y}-\mu)} |\Sigma|^{-\frac{\nu_0+d+2}{2}} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)}$
$= |\Sigma|^{-\frac{n+\nu_0+d+2}{2}} e^{-\frac{1}{2}\text{tr}[\underbrace{\sum(y-\mu)(y-\mu)^T}_{\text{S}}\Sigma^{-1}]} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1})} e^{-\frac{\kappa_0}{2n}\text{tr}[\underbrace{\sum(\mu-\mu_0)(\mu-\mu_0)^T}_{\text{M}}\Sigma^{-1}]}$
$= |\Sigma|^{-\frac{n+\nu_0+d+2}{2}} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1})} e^{-\frac{1}{2}\text{tr}[(S+\frac{k_0}{n}M)\Sigma^{-1}]}$

Now, $\text{tr}[(S + \frac{k_0}{n}M)\Sigma^{-1}] \propto (1 + \frac{k_0}{n})[\sum \mu\mu^T - \sum \frac{(y_i + \frac{k_0}{n}\mu_0)}{1+\frac{k_0}{n}}\mu^T + ..]\Sigma^{-1}$

Therefore, $k_n = k_0 + n$, $\nu_n = \nu_0 + n$, $\mu_n = \frac{k_0\mu_0 + \sum y_i}{k_0+n}$ and $\Lambda_n = S + \Lambda_0$

Exercise 2.9:

$y|X, \beta \sim N(X\beta, (w\Lambda)^{-1})$
$\beta \sim N(\mu, (wK)^{-1})$
$w \sim \text{Gamma}(a, b)$

Conditional posterior:
$p(\beta|w, y_1, ..., y_n) \propto p(y_1, .., y_n|X, \beta, w)p(\beta|w)$
$p(y_1, .., y_n|X, \beta, w) \propto \exp\Big[-\frac{1}{2}(y - X\beta)^T(w\Lambda)(y - X\beta)\Big]$
$p(\beta|w) \propto \exp\Big[-\frac{1}{2}(\beta - \mu)^T(wK)(\beta - \mu)\Big]$

Now, $(y - X\beta)^T(w\Lambda)(y - X\beta) = (y - X\hat{\beta} - X\beta + X\hat{\beta})^T(w\Lambda)(y - X\hat{\beta} - X\beta + X\hat{\beta})$
$\propto (\beta - \hat{\beta})^T(wX^T\Lambda X)(\beta - \hat{\beta})$
After some manipulation we get:

$\beta|w, y_1, ..., y_n \sim N((X^T\Lambda X + K)^{-1}(X^T\Lambda My + K\mu), (w(X^T\Lambda X + K))^{-1})$ where $M = X(X^TX)^{-1}X^T$

Exercise 2.10:

$p(w|y_1, ..., y_n) \propto \int p(y_1, .., y_n|X, \beta, w)p(\beta|w)p(w)d\beta$.
Using the result from 2.9 and after some manipulation, we get:
$w|y_1, .., y_n \sim \text{Gamma}(a^*, b^*)$
where $a^* = a+n/2$ and $b^* = b+(y^T\Lambda y+\mu^T K\mu-(X^T\Lambda My+K\mu)^T(X^T\Lambda X+K)^{-1}(X^T\Lambda My+K\mu))/2$

Exercise 2.11:

Full conditional of $w|\beta, y \sim \text{Gamma}(a_1, b_1)$, where $a_1 = a + (n + p)/2$ and $b_1 = b + \frac{1}{2}(y - X\beta)^T\Lambda(y - X\beta) + \frac{1}{2}(\beta - \mu)^T K(\beta - \mu)$.
If we integrate out w, we get the marginal posterior of $\beta|y$:
$p(\beta|y) \propto (b + \frac{1}{2}(y - X\beta)^T\Lambda(y - X\beta) + \frac{1}{2}(\beta - \mu)^T K(\beta - \mu))^{a_1}$ which is actually a non-central t distribution.

Exercise 2.12:

Response: Distance
Predictors: Age, Sex
Here we have assumed that the predictors are categorical. Thus, for Age, 8 years is the reference group and for Sex, Female is the reference group. Having said that, we get 5 coefficients including the intercept. So after running the Gibbs sampler (initial values: $\mu = 0$, $a = 2$ and $b = 1$) we get the estimates of β as follows:

parameter	Estimate from bayesian model	Estimate from LS
Intercept	20.14	20.81
Age of 10 years	1.37	0.98
Age of 12 years	2.82	2.46
Age of 14 years	4.21	3.91
Gender	2.68	2.32

Results are not exactly similar but not very far from each other. In a fixed-effects model the shapes of projections of deviance contours onto pairs of fixed-effects parameters are consistent. For mixed models these traces can fail to be linear.

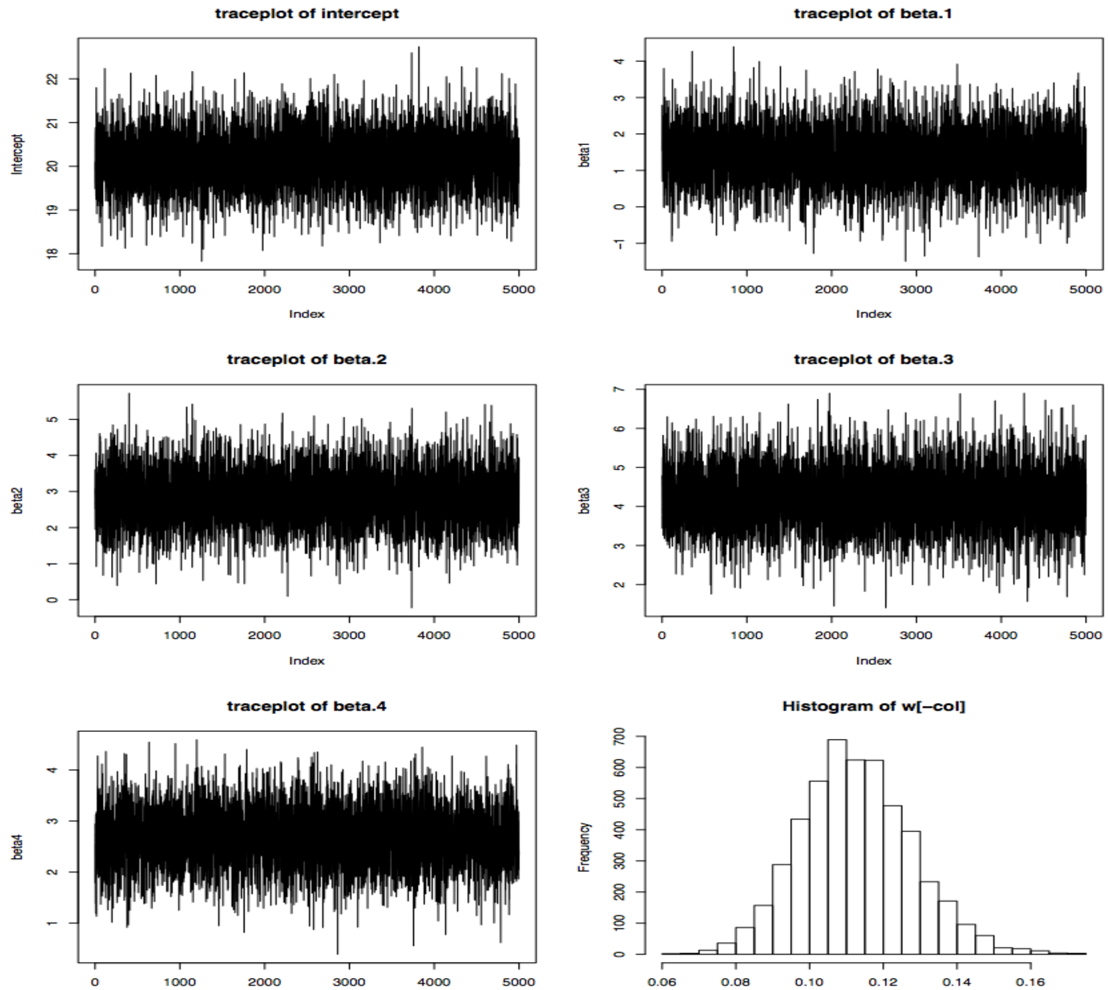

Figure 1: 2.11 traceplots and histogram

Exercise 2.13:

$p(\lambda_i|y,\beta,w) \propto p(y_i|\beta,w,\lambda_i)p(\lambda_i) = \sqrt{\frac{w\lambda_i}{2\pi}} \exp\{-\frac{w\lambda_i}{2}(y_i - x_i^T\beta)^2\}\frac{\tau^\tau}{\Gamma\tau}\lambda_i^{\tau-1}\exp\{-\tau\lambda_i\}$
$= \lambda_i^{\tau+\frac{1}{2}-1}\exp\{-\lambda_i(\tau + \frac{w}{2}(y_i - x_i^T\beta)^2)\}$

$\lambda_i \sim \text{Gamma}(\tau + \frac{1}{2}, \tau + \frac{w}{2}(y_i - x_i^T\beta)^2)$

Exercise 2.14:

After running a Gibbs sampler, we get the estimates as:

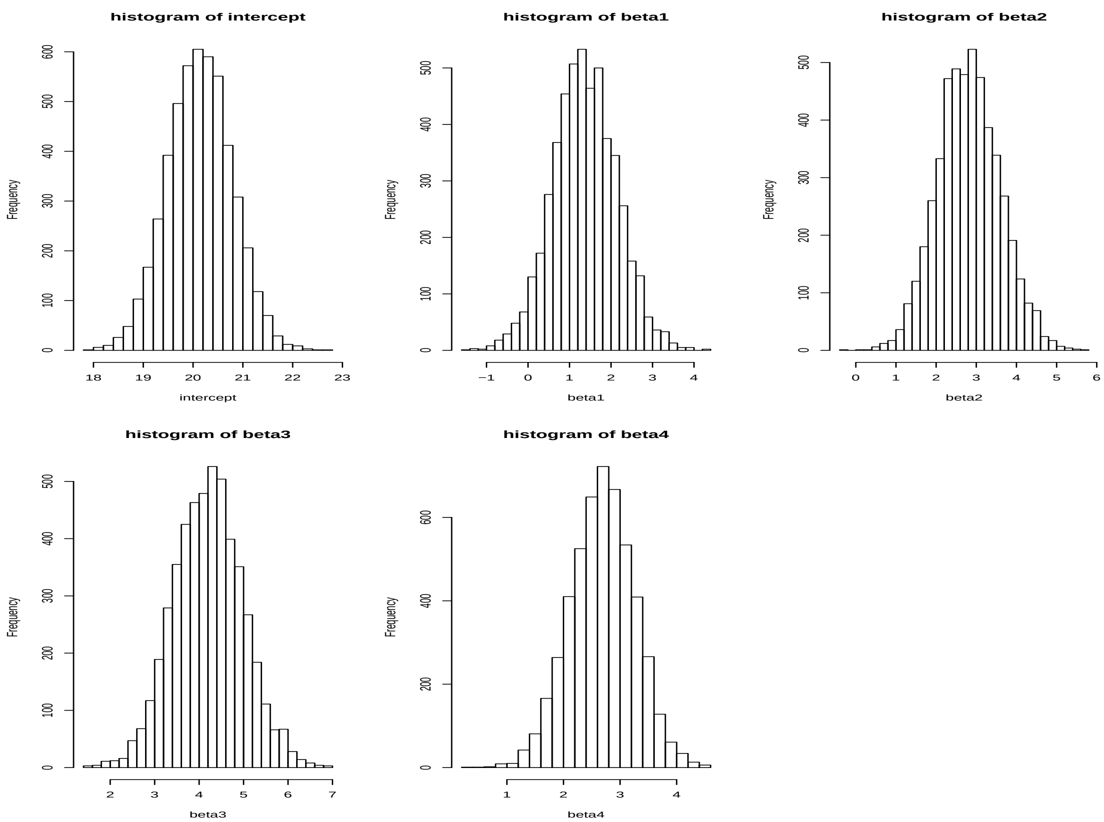

Figure 2: 2.11 histogram of β

parameter	Estimates
Intercept	13.93
Age of 10 years	0.91
Age of 12 years	1.97
Age of 14 years	2.96
Gender	1.69

We run the Gibbs sampler for 5000 iterations.

Exercise 2.15:

It seems from the two fits that the results differ from each other. The difference between the assumptions of the first one and the second one is that the matrix Λ is not constant any more. The data provided for this section is a longitudinal data consisting of repeated measurements on the same subject taken over time. Thus our goal is to characterize the time trends within subjects and between subjects. In this situation, we may include the indicator of the subject on which the measurement has been made to modify the model.