MAS8303 Modern Bayesian Inference
Solutions to Exercises 2.5

University of Newcastle upon Tyne

Semester 1, 2012-3

1. Observations are made on the numbers of caterpillars on commercially grown cabbages in J plots. The number of observations in plot j is n_{ij}. Let the number of caterpillars on the i^{th} cabbage in plot j be Y_{ij}. Given the values of $\lambda_1, \dots, \lambda_J$, we have

$$Y_{ij} \mid \lambda_j \sim \text{Poisson}(\lambda_j),$$

a Poisson distribution with mean λ_j, and $Y_{11}, \dots, Y_{n_J J}$ are conditionally independent.

Let $\eta_j = \log(\lambda_j)$. Given the values of μ and τ, we have

$$\eta_j \mid \mu, \tau \sim N(\mu,\ \tau^{-1}),$$

a normal distribution with mean μ and precision τ, and $\eta_1, \dots, \eta_J$ are conditionally independent.

We have independent prior distributions for μ and τ with $\mu \sim N(m, v)$ and $\tau \sim \text{gamma}(a, b)$.

We make observations $Y_{ij} = y_{ij}$ and wish to use a Gibbs sampler to evaluate the posterior distribution.

Find a function proportional to the density of the full conditional distribution of η_j.

Solution

In plot j we have

$$Y_{ij} \mid \lambda_j \sim \text{Poisson}(\lambda_j),$$

for $i = 1, \dots, n$. So

$$\Pr(Y_{1j} = y_{1j}, \dots, Y_{n_j j} = y_{n_j j} \mid \lambda_j) = \prod_{i-1}^{n_j} \frac{e^{-\lambda_j} \lambda_j \lambda_j y_{ij}}{y_{ij}!} \propto e^{-n_j \lambda_j} \lambda_j^{s_j}$$

where

$$s_j = \sum_{i=1}^{n_j} y_{ij}.$$

However $\lambda_j = e^{\eta_j}$ so

$$\Pr(Y_{1j} = y_{1j}, \dots, Y_{n_j j} = y_{n_j j} \mid \eta_j) \propto \exp\{-n_j e^{\eta_j}\} e^{\eta_j s_j} = \exp\{\eta_j s_j - n_j e^{\eta_j}\}.$$

The pdf of $\eta_j \mid \mu, \tau$ is

$$(2\pi)^{-1/2} \tau^{1/2} \exp\left\{-\frac{\tau}{2}(\eta_j - \mu)^2\right\} \propto \exp\left\{-\frac{\tau}{2}(\eta_j - \mu)^2\right\}.$$

Hence the fcd density of η_j is proportional to

$$\exp\left\{-\frac{\tau}{2}(\eta_j-\mu)^2\right\}\exp\{\eta_j s_j - n_j e^{\eta_j}\} = \exp\left\{\eta_j s_j - n_j e^{\eta_j} - \frac{\tau}{2}(\eta_j-\mu)^2\right\}.$$

2. A particular surgical operation performed on patients with a serious condition is hazardous and a proportion of the patients die during surgery. Researchers wish to investigate the relationship between the death rate and the age of the patient. We have the following model. Let θ_x be the death rate for patients aged x years. That is, given θ_x, the probability of death is θ_x. Let

$$\eta_x = \log\left(\frac{\theta_x}{1-\theta_x}\right).$$

We suppose that

$$\eta_x = a + bx$$

for some unknown parameters $a,\ b$.

We develop a prior distribution as follows. Consider two ages, $x = 50$ and $x = 70$. Our marginal prior distributions for η_{50} and η_{70} are $\eta_{50} \sim N(m_{50}, v_{50})$ and $\eta_{70} \sim N(m_{70}, v_{70})$. The prior correlation of η_{50} and η_{70} is 0.8. We assess

$$\Pr(\theta_{50} < 0.05) = \Pr(\theta_{50} > 0.20) = \Pr(\theta_{70} < 0.1) = \Pr(\theta_{70} > 0.4) = 0.025.$$

(a) Find the values of $m_{50},\ v_{50},\ m_{70},\ v_{70}$ and the covariance of η_{50}, η_{70}.

(b) Find the joint prior distribution of a, b.

<u>Solution</u>

(a) We have

$$\Pr(\theta_{50} < 0.05) = \Pr(\theta_{50} > 0.20) = \Pr(\theta_{70} < 0.1) = \Pr(\theta_{70} > 0.4) = 0.025$$

so

$$\Pr(\eta_{50} < -2.944) = \Pr(\eta_{50} > -1.386) = \Pr(\eta_{70} < -2.197) = \Pr(\eta_{70} > -0.405) = 0.025$$

since, for example,

$$\log\left(\frac{0.05}{1-0.05}\right) = -2.944.$$

The 0.975 point of $N(0,1)$ is 1.96 so

$$\begin{aligned} m_{50} - 1.96\sqrt{v_{50}} &= -2.944, \\ m_{50} + 1.96\sqrt{v_{50}} &= -1.386. \end{aligned}$$

Hence

$$\begin{aligned} m_{50} &= \frac{-2.944-1.386}{2} = \underline{-2.165}, \\ \sqrt{v_{50}} &= \frac{-1.386-(-2.944)}{2\times 1.96} = 0.3974, \\ v_{50} &= \underline{0.1580}. \end{aligned}$$

Similarly

$$\begin{aligned} m_{70} - 1.96\sqrt{v_{70}} &= -2.197, \\ m_{70} + 1.96\sqrt{v_{70}} &= -0.405. \end{aligned}$$

Hence

$$\begin{aligned} m_{70} &= \frac{-2.197 - 0.405}{2} = \underline{-1.301}, \\ \sqrt{v_{70}} &= \frac{-0.405 - (-2.197)}{2 \times 1.96} = 0.4571, \\ v_{70} &= \underline{0.2090}. \end{aligned}$$

Now

$$\frac{\text{Covar}(\eta_{50}, \eta_{70})}{\sqrt{v_{50}v_{70}}} = 0.8.$$

Hence

$$\text{Covar}(\eta_{50}, \eta_{70}) = 0.8\sqrt{v_{50}v_{70}} = 0.8 \times 0.3974 \times 0.4571 = \underline{0.1454}.$$

(b) We have

$$\begin{aligned} \eta_{50} &= a + 50b, \\ \eta_{70} &= a + 70b. \end{aligned}$$

Hence

$$\begin{aligned} b &= \frac{\eta_{70} - \eta_{50}}{20}, \\ a &= \eta_{50} - 50b = \frac{20\eta_{50} - 50\eta_{70} + 50\eta_{50}}{20} = \frac{70\eta_{50} - 50\eta_{70}}{20}. \end{aligned}$$

Hence

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 70 & -50 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \eta_{50} \\ \eta_{70} \end{pmatrix}.$$

If $(a, b)^T = M(\eta_{50}, \eta_{70})^T$ and the mean of $(\eta_{50}, \eta_{70})^T$ is $\underline{m}$ and the variance of $(\eta_{50}, \eta_{70})^T$ is V then the mean of $(a, b)^T$ is $M\underline{m}$ and the variance of $(a, b)^T$ is MVM^T.
The distribution of $(a, b)^T$ is bivariate normal with mean vector

$$\text{E}\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 70 & -50 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -2.165 \\ -1.301 \end{pmatrix} = \begin{pmatrix} -4.3250 \\ 0.0432 \end{pmatrix}$$

and covariance matrix

$$\begin{aligned} \text{Var}\begin{pmatrix} a \\ b \end{pmatrix} &= \frac{1}{400}\begin{pmatrix} 70 & -50 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 0.1580 & 0.1454 \\ 0.1454 & 0.2090 \end{pmatrix}\begin{pmatrix} 70 & -1 \\ -50 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0.69725 & -0.010155 \\ -0.010155 & 0.0001905 \end{pmatrix}. \end{aligned}$$

3. In an experiment on student learning, randomly selected students are assigned to groups which are given different amounts of tuition. Suppose group i has n_i students who are given $t_i + 30$ hours of tuition.

   At the end of the experiment the students are given a test. Suppose that a student's percentage mark is Z. Let $X = \ln(Z)$. Suppose that, for a student in group i, we assume $X \sim N(\alpha + \beta t_i, \sigma^2)$, where $\sigma = 0.1$. Instead of the actual percentage marks, all that is recorded is whether each student passes or fails the test. A student passes if $Z \geq 40$, that is $X \geq \ln 40$.

   Let y_i be the number of students in group i who pass the test.

(a) Express this model as a generalised linear model.

(b) State the link function and error function.

(c) Find the linear predictor.

(d) Use BRugs to evaluate the posterior distributions of α and β. You may use independent priors for α and β with

$$\alpha^* \sim N(0.1, 0.01), \quad \alpha = \alpha^* + \ln 40 \quad \text{and} \quad \beta \sim N(0.0, 0.0004).$$

The data are as follows.

t_i	n_i	y_i
-10	30	19
0	40	30
10	30	27

(e) What happens if we do not assume that $\sigma = 0.1$ but allow σ^2 to be unknown?

<u>**Solution**</u>

(In an assignment, you should start by giving the background to the problem, the model and the prior as given in the question).

Given the model parameters, the probability of passing with $t_i + 30$ hours of tuition is

$$\begin{aligned}
\Pr(Z \geq 40) = \Pr(X \geq \ln 40) &= \Pr\left(\frac{X - \alpha - \beta t_i}{\sigma} \geq \frac{\ln 40 - \alpha - \beta t_i}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{\ln 40 - \alpha - \beta t_i}{\sigma}\right) \\
&= \Phi\left(\frac{\alpha + \beta t_i - \ln 40}{\sigma}\right)
\end{aligned}$$

where $\Phi()$ is the standard normal distribution function.

Thus our model is a generalised linear model.

- The error distribution is binomial.
- The linear predictor is
$$\eta_i = \frac{\alpha}{\sigma} + \frac{\beta}{\sigma} t_i - \frac{\ln 40}{\sigma}.$$
- The link function is the probit link with $\eta_i = \Phi^{-1}(p_i)$ where p_i is the probability of a pass with $t_i + 30$ hours of tution and $\Phi^{-1}()$ is the inverse of $\Phi()$.
- Note that the known term $\ln(40)/\sigma$ in the linear predictor is known as an *offset*.

We can evaluate the posterior distribution using the BRugs software to implement a Gibbs sampler. The following is a suitable model specification.

```
model tuition

{
 for (i in 1:3)
   {y[i]~dbin(p[i],n[i])
    probit(p[i])<-10*(alphastar+beta*t[i])
    }

 alphastar~dnorm(0.1, 100)
 alpha<-alphastar+log(40)
 beta~dnorm(0, 2500)
 }
```

Convergence of the sampler was checked by using two parallel chains, using the commands in the following session listing.

```
> modelCheck("tuitionbug.txt")
model is syntactically correct
> modelData("tuitiondata.txt")
data loaded
> modelCompile(2)
model compiled
> modelInits("inits1.txt")
Initializing chain 1: model is initialized
> modelInits("inits2.txt")
Initializing chain 2: model is initialized
> samplesSet(c("alpha","beta"))
monitor set for variable 'alpha'
monitor set for variable 'beta'
> modelUpdate(1000)
```

The two initial value files, `inits1.txt` and `inits2.txt` were as follows.

"inits1":

```
list(alphastar=0.15,beta=0.00)
```

"inits2":

```
list(alphastar=0.02,beta=0.01)
```

Figure 1 shows plots of sampled values against iteration number for a burn-in of 1000 samples for both α and β for the two chains. Chain 1 is shown as a blue line. Chain 2 is shown as a red line. Although mixing is a little slow, suggesting that a large number of samples should be collected, there seems to be no problem with convergence. We could try more widely separated initial values.

To compute the posterior distribution, 20000 samples were collected from a single chain after a burn-in of 1000 iterations. These gave the following summaries of the posterior distribution.

```
          mean       sd  MC_error  val2.5pc   median val97.5pc start sample
alpha 3.765000 0.014200 2.251e-04 3.7380000 3.765000  3.794000  1001  20000
beta  0.004568 0.001881 3.855e-05 0.0009511 0.004531  0.008238  1001  20000
```

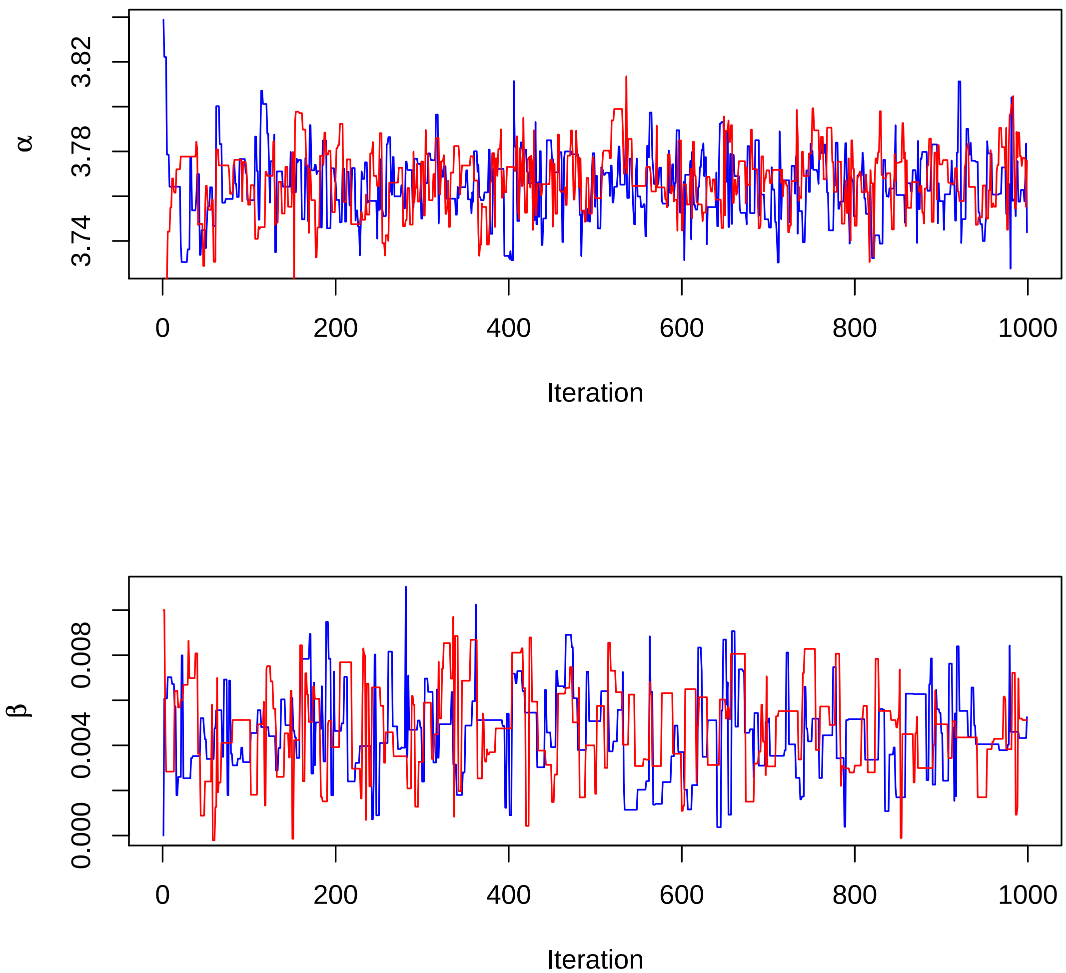

Figure 1: History plots for α and β in a burn-in of 1000 samples.

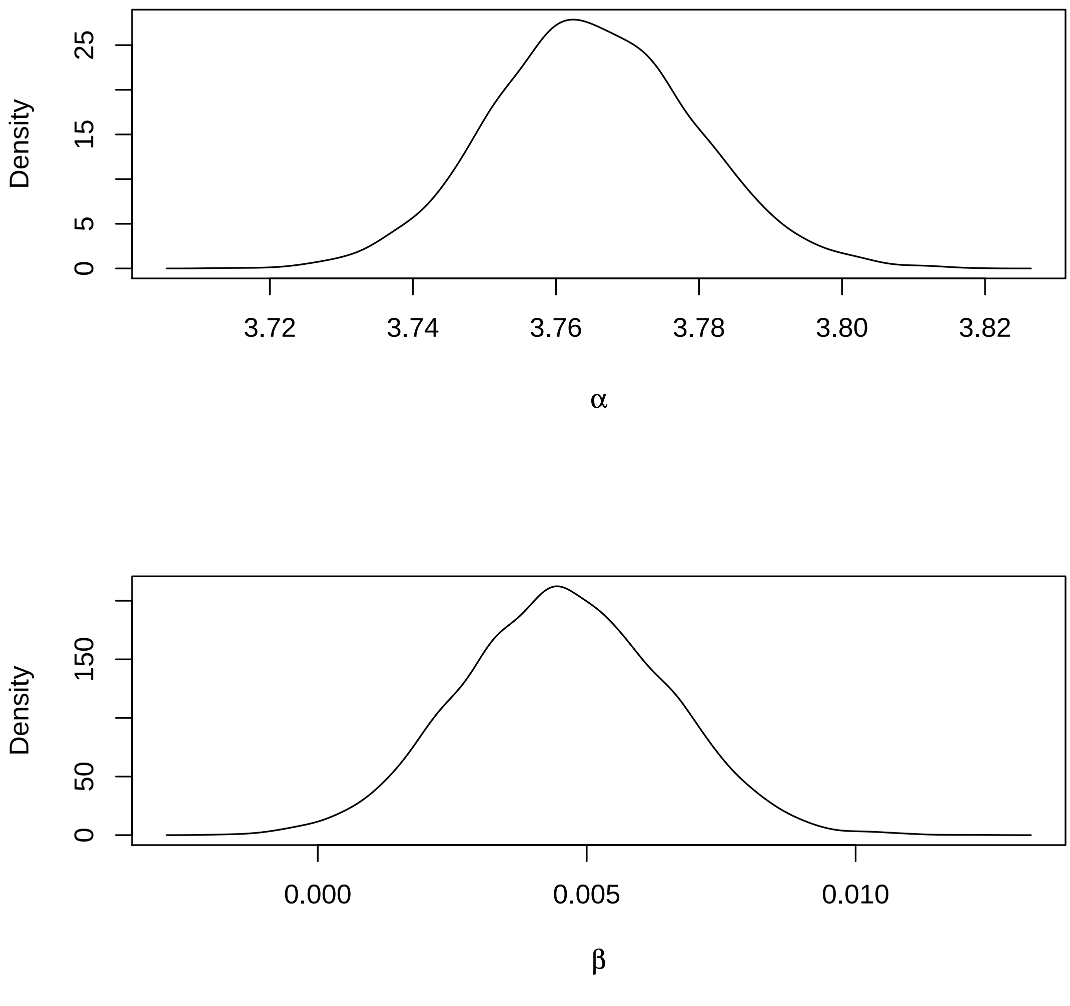

Figure 2: Posterior marginal densities of α and β.

Thus, for α, we have a posterior mean of 3.765 and a posterior standard deviation of 0.0142, compared to the prior mean of $0.1 + \ln 40 = 3.789$ and prior standard deviation of 0.1. For β, we have a posterior mean of 0.00457 and a posterior standard deviation of 0.00188, compared to the prior mean of 0.0 and prior standard deviation of 0.02.

Figure 2 shows the posterior marginal densities of α and β. These were obtained by first using `samplesSample` to extract the sampled values of the parameters and then using the following R commands.

```
> adens<-density(alpha,adjust=1.5)
> bdens<-density(beta,adjust=1.5)
> pdf("densitiesex2.pdf",height=7)
> par(mfrow=c(2,1))
> plot(adens$x,adens$y,type="l",xlab=expression(alpha),ylab="Density")
> plot(bdens$x,bdens$y,type="l",xlab=expression(beta),ylab="Density")
> par(mfrow=c(1,1))
> dev.off()
```

There is evidence that increasing tuition time increases the probability of passing since most of the probability for β is for values where $\beta > 0$.

Recall that the linear predictor is

$$\eta_i = \frac{\alpha - \ln 40}{\sigma} + \frac{\beta}{\sigma} t_i.$$

Suppose that we replace σ by some other value $\tilde{\sigma}$. Then, by replacing α by

$$\tilde{\alpha} = \frac{\tilde{\sigma}}{\sigma}(\alpha - \ln 40) + \ln 40 \qquad \text{so} \qquad \frac{\tilde{\alpha} - \ln 40}{\tilde{\sigma}} = \frac{\alpha - \ln 40}{\sigma}$$

and replacing β by

$$\tilde{\beta} = \frac{\tilde{\sigma}}{\sigma}\beta \qquad \text{so} \qquad \frac{\tilde{\beta}}{\tilde{\sigma}} = \frac{\beta}{\sigma},$$

we obtain the same value of η_i as before. Therefore σ is not identified by the data and there is no point in treating it as an unknown parameter.