Applied Bayesian Inference Assignments October 4-8, 2010

1. **[6 marks] Bayes' Theorem and The Prisoners' Dilemma**

   a) **[3m]** Consider three prisoners A, B, and C, exactly one of whom will be pardoned; the other two will be executed. Let A, B, and C, respectively, denote the events that prisoner A, B, or C will be pardoned. In the absence fo any other evidence to the contrary, it is reasonable to assume the 'prior probabilities'

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

   The warden now enters prisoner A's cell and tells him that B will be executed, that is, the event $W(A, B)$ occurs. Prisoner A thinks that he now possesses an increased probability of 0.5 of being pardoned. Using Bayes' theorem, calculate $P(A|W(A, B))$ under the assumptions

$$\begin{aligned} P(W(A,B)|A) &= P(W(A,C)|A) = \frac{1}{2}, \\ P(W(A,C)|B) &= 1, \\ P(W(A,B)|C) &= 1. \end{aligned}$$

   Is that equal to 0.5? Prisoner C is listening through a hole in the wall. What is his updated probability of execution?

   Answer:

$$\begin{aligned} P(A|W(A,B)) &= \frac{P(W(A,B)|A)P(A)}{P(W(A,B)|A)P(A) + P(W(A,B)|B)P(B) + P(W(A,B)|C)P(C)} \\ &= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3}} = \frac{1}{3} \\ P(C|W(A,B)) &= \frac{P(W(A,B)|C)P(C)}{P(W(A,B)|A)P(A) + P(W(A,B)|B)P(B) + P(W(A,B)|C)P(C)} \\ &= \frac{2}{3} \end{aligned}$$

   So, the information provided by the warden does not influence prisoner A's information. But the updated probability of execution for prisoner C is now halved to $\frac{1}{3}$.

2. **[10 marks] Likelihood function and ML estimation**
The file `ozone.txt` gives measurements of ozone partial pressure, y_i, in millibars in each of 15 atmospheric layers where each layer, x_i, is approximately 2 km in height. The layers have been scaled for convenience from -7 to +7.

   (a) **[1m]** Assuming the linear model $y_i|\beta_0, \beta_1 = \beta_0 + \beta_1 x_i + e_i$ for $i = 1, \ldots, n$ where $e_i \overset{iid}{\sim} N(0, \sigma^2)$ with known σ^2, write down the likelihood function of the unknown parameters β_0, β_1.

   Answer:
   As the errors $e_i \overset{iid}{\sim} N(0, \sigma^2)$, the observations y_i are conditionally independent, given β_0, β_1. Thus

$$\begin{aligned} f(y_1, \ldots, y_n|\beta_0, \beta_1) &\overset{ind.}{=} \prod_{i=1}^{n} f(y_i|\beta_0, \beta_1) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\} \\ &= (2\pi)^{-n/2}\sigma^{-n} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\} \end{aligned}$$

   (b) **[4m]** Find the ML estimates of β_0, β_1 and calculate the Fisher information matrix. (Do not yet plug in the actual data but give the general formulae).

   Answer:
   Calculate log-likelihood, differentiate and set to zero:

$$\begin{aligned} l(\beta_0, \beta_1) &= \log f(y_1, \ldots, y_n|\beta_0, \beta_1) = c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \\ \frac{\partial}{\partial\beta_0} l(\beta_0, \beta_1) &= (-1)\frac{-2}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \\ &= \frac{1}{\sigma^2}(n\bar{y} - n\beta_0 - \beta_1 n\bar{x}) \\ &= 0 \qquad \Longleftrightarrow \qquad \beta_0 = \bar{y} - \beta_1\bar{x} \\ \frac{\partial}{\partial\beta_1} l(\beta_0, \beta_1) &= (-1)\frac{-2}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i \\ &= \frac{1}{\sigma^2}\left(\sum x_i y_i - \beta_0 n\bar{x} - \beta_1 \sum x_i^2\right) \\ &= \frac{1}{\sigma^2}\left(\sum x_i y_i (\bar{y} - \beta_1\bar{x}) n\bar{x} - \beta_1 \sum x_i^2\right) \\ &= \frac{1}{\sigma^2}\left(\sum (x_i - \bar{x})(y_i - \bar{y}) - \beta_1 \sum (x_i - \bar{x})^2\right) \\ &= 0 \qquad \Longrightarrow \qquad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \end{aligned}$$

   Now calculate matrix of 2. derivatives for the Fisher information matrix:

$$\begin{aligned} \frac{\partial^2 l}{\partial\beta_0^2} &= -\frac{n}{\sigma^2} \\ \frac{\partial^2 l}{\partial\beta_0\beta_1} &= -\frac{n\bar{x}}{\sigma^2} \end{aligned}$$

$$\frac{\partial^2 l}{\partial \beta_1^2} = -\frac{\sum x_i^2}{\sigma^2}$$

$$\begin{aligned} I(\beta_0, \beta_1) &= E\left(-\frac{\partial^2}{\partial \beta^2} \log l(\beta_0, \beta_1)\right) \\ &= E\begin{bmatrix} \frac{n}{\sigma^2} & \frac{n\bar{x}}{\sigma^2} \\ \frac{n\bar{x}}{\sigma^2} & \frac{\sum x_i^2}{\sigma^2} \end{bmatrix} \end{aligned}$$

The (asymptotic) covariance matrix of the ML estimates is given by the inverse of $I(\beta_0, \beta_1)$, i.e.

$$I(\beta_0, \beta_1)^{-1} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \begin{bmatrix} \frac{1}{n}\sum x_i^2 & \bar{x} \\ \bar{x} & 1 \end{bmatrix}$$

Thus, the square roots of the diagonal elements give s.e. estimates

$$\begin{aligned} se(\hat{\beta}_0) &= \sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2} \frac{1}{n} \sum x_i^2} \\ se(\hat{\beta}_1) &= \sqrt{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}} \end{aligned}$$

As σ^2 is unknown in practice, an estimate is substituted for σ^2.

(c) **[2m]** Show that the Least Squares estimates of β_0, β_1 are the same as the ML estimates.

Answer:

According to the least squares principle, the parameters are estimated by minimizing the sum of squared deviations of the observations from the model, i.e. $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$. Under the normal distribution assumption, this is exactly the minimization problem solved by the ML estimates.

(d) **[1m]** Using the function `lm` in R or Splus and now assuming more realistically that σ^2 is unknown, give LS estimates of the parameters for

1. the linear model $y_i = \beta_0 + \beta_1 x_i + e_i$,
2. the quadratic model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$ and
3. for the cubic model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + e_i$.

Answer:

```
ozone <- read.table("ozone.txt", header=TRUE)
oz <- data.frame(ozone)
attach(oz)

x2 <- x^2
x3 <- x^3

oz.lm <- lm(y~x)
oz.qm <- lm(y~x+x2)
oz.cm <- lm(y~x+x2+x3)
```

```
> summary(oz.lm)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-43.683 -10.936   1.835  13.708  25.200

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  78.1731     2.2395  34.907   <2e-16 ***
x             0.5729     0.4879   1.174    0.245
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.78 on 59 degrees of freedom
Multiple R-Squared: 0.02283,    Adjusted R-squared: 0.00627
F-statistic: 1.379 on 1 and 59 DF,  p-value: 0.2451

> summary(oz.qm)

Call:
lm(formula = y ~ x + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-6.7382 -1.9054  0.1729  1.8438  6.8869

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 98.00543    0.63954  153.24  < 2e-16 ***
x           -0.70418    0.09568   -7.36 7.29e-10 ***
x2          -1.01997    0.02503  -40.75  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.109 on 58 degrees of freedom
Multiple R-Squared: 0.967,      Adjusted R-squared: 0.9659
F-statistic: 850.4 on 2 and 58 DF,  p-value: < 2.2e-16

> summary(oz.cm)

Call:
lm(formula = y ~ x + x2 + x3)

Residuals:
     Min       1Q   Median       3Q      Max
-6.91206 -1.59239 -0.08232  1.51601  5.46965
```

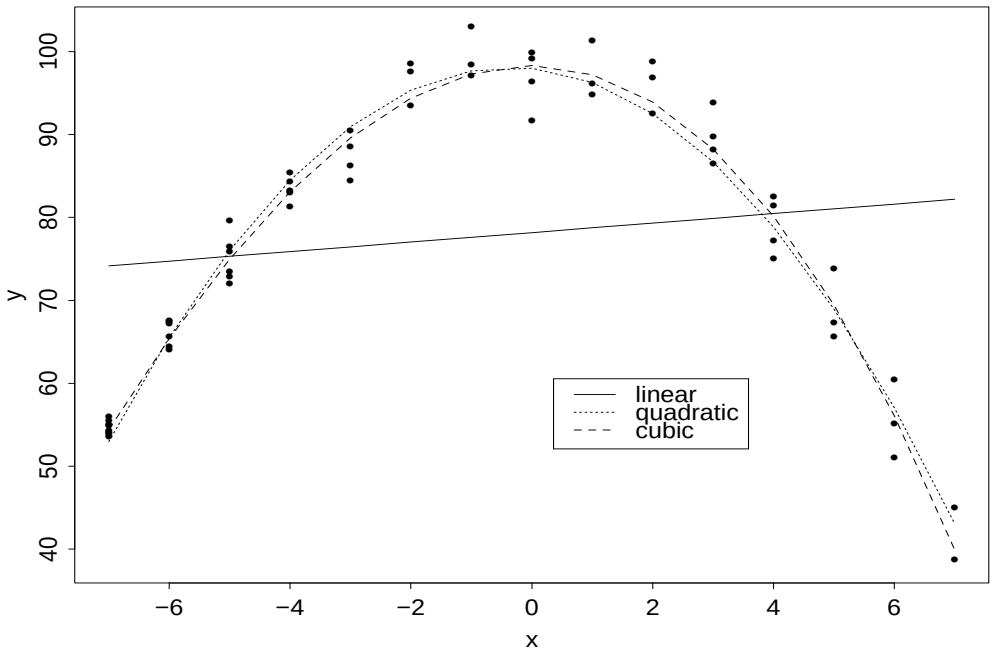

Figure 1: Ozone data and fitted linear, quadratic and cubic models.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 98.312063   0.596123 164.919  < 2e-16 ***
x           -0.041489   0.215553  -0.192  0.84805
x2          -1.043527   0.024093 -43.312  < 2e-16 ***
x3          -0.020324   0.006033  -3.369  0.00136 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.864 on 57 degrees of freedom
Multiple R-Squared: 0.9725,     Adjusted R-squared: 0.9711
F-statistic: 671.9 on 3 and 57 DF,  p-value: < 2.2e-16
```

(e) **[1m]** Compare the models with the data by plotting the model fits on the same graph as your data.

Answer: See Figure 7.

(f) **[1m]** Using the R function `AIC()`, calculate Akaike's and Schwarz' information criteria for the three models and interpret the results.

Answer:

```
> #Akaike's information criterion
> AIC(oz.lm,oz.qm,oz.cm,k=2)
      df      AIC
```

```
oz.lm  3 521.1404
oz.qm  4 316.4197
oz.cm  5 307.3431
> #Schwarz or Bayesian information criterion (BIC)
> AIC(oz.lm,oz.qm,oz.cm,k=log(nrow(ozone)/2/pi))
      df      AIC
oz.lm  3 521.9594
oz.qm  4 317.5117
oz.cm  5 308.7081
```

Both AIC and BIC are smallest for the cubic regression model, thus this is the preferred model.

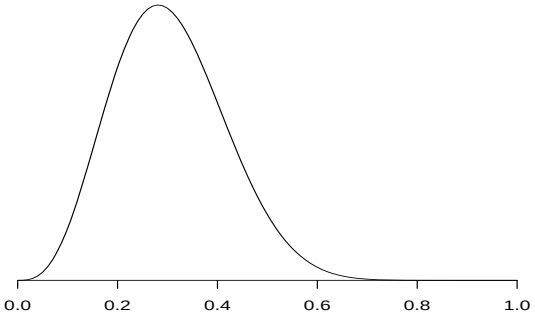

Figure 2: Posterior density of θ.

3. **[4 marks] Beta-Binomial**
Suppose there is a Beta(4,4) prior distribution on the probability θ that a coin will yield a *head* when spun in a specified manner. The coin is independently spun ten times, and *heads* appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3. Calculate your exact posterior density (up to a proportionality constant) for θ and plot it using R or Splus.

Answer:

Prior pdf:

$$f(\theta) \propto \theta^3(1-\theta)^3$$

Likelihood:

$$\begin{aligned} P(\text{data}|\theta) &= \binom{10}{0}(1-\theta)^{10} + \binom{10}{1}\theta(1-\theta)^9 + \binom{10}{2}\theta^2(1-\theta)^8 \\ &= (1-\theta)^{10} + 10\theta(1-\theta)^9 + 45\theta^2(1-\theta)^8 \end{aligned}$$

Posterior:

$$\begin{aligned} f(\theta|\text{data}) &\propto P(\text{data}|\theta)f(\theta) \\ &\propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}. \end{aligned}$$

R-code:

```
theta <- seq(0,1,0.01)
dens  <- theta^3*(1-theta)^13 + 10*theta^4*(1-theta)^12 +
         45*theta^5*(1-theta)^11
plot(theta,dens,ylim=c(0,1.1*max(dens)),type="l",xlab="theta",ylab="",
     xaxs="i",yaxs="i",yaxt="n",bty="n",cex=1)
```

4. **[3 marks] Prior Predictive Distribution**
   Assume $X|\theta \sim \text{Binomial}(n, \theta)$ and a uniform prior distribution for θ. What is the prior predictive pdf of X?

Answer:
$X|\theta \sim$ Binomial(n, θ) and $\theta \sim$ Beta(1,1). The prior predictive pdf of X is

$$\begin{aligned} f(x) &= \int_0^1 f(x|\theta) f(\theta) d\theta \\ &= \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta \\ &= \binom{n}{x} \frac{\Gamma(1+x)\Gamma(1+n-x)}{\Gamma(n+2)} \\ &= \frac{n!}{x!(n-x)!} \frac{x!(n-x)!}{(n+1)!} \\ &= \frac{1}{n+1} \qquad \text{for } x = 0, 1, \ldots, n. \end{aligned}$$

5. **[3 marks]**
   Laplace (1774) claimed that the probability that an event which has occured n times, and has hitherto not failed, will occur again is $(n+1)/(n+2)$. Suggest grounds for this assertion.

   Answer:

   Let θ be the probability that the event occurs and assume observations are exchangeable. $X_i|\theta \sim$ Bernoulli(θ) for $i = 1, \ldots, n$ and assume a uniform distribution for θ. Then the posterior distribution of θ is Beta$(1+n, 1)$.

   As shown in class, the posterior predictive probability for success is the posterior mean of θ. As $\theta \sim$ Beta$(1+n, 1)$, this is $\frac{n+1}{n+2}$.

6. **[15 marks] Binomial Example**
In a research program on human health risk from recreational contact with water contaminated with pathogenic microbiological material, the National Institute of Water and Atmosphere (NIWA) instituted a study to determine the quality of NZ stream water at a variety of catchment types. This study is documented in McBride et al. (2002) where $n = 116$ one-liter water samples from sites identified as having a heavy environmental impact from birds (seagulls) and waterfowl. Out of these samples, $x = 17$ samples contained Giardia cysts. Let θ denote the true probability that a one-liter water sample from this type of site contains Giardia cysts.

   (a) **[1m]** What is the conditional distribution of X, the number of samples containing Giardia cysts, given θ?

   Answer:
   $X|\theta \sim Binomial(n = 116, \theta)$

   (b) **[2m]** Before the experiment, the NIWA scientists elicited that the expected value of θ is 0.2 with a standard deviation of 0.16. Determine the parameters α and β of a Beta prior distribution for θ with this prior mean and standard deviation. (Round α and β to the nearest integer).

   Answer:
   If $\theta \sim Beta(\alpha, \beta)$, then

$$\begin{aligned} E[\theta] &= \frac{\alpha}{\alpha+\beta} \stackrel{!}{=} 0.2 \\ Var(\theta) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \stackrel{!}{=} 0.16^2. \end{aligned}$$

   Solving these two equations with respect to α and β yields $\alpha = 1$ and $\beta = 4$.

   (c) **[1m]** Find the posterior distribution of θ and summarize it by its posterior mean and standard deviation.

   Answer:
   $\theta|x = 17 \sim Beta(18, 103)$. Thus $E[\theta|x = 17] = \frac{18}{18+103} = 0.149$, $Var(\theta|x = 17) = \frac{18\times103}{(18+103)^2(18+103+1)} = 0.001037957$ and $sd(\theta|x = 17) = 0.0322$.

   (d) **[1m]** Verify that the posterior mean is a weighted average of prior mean and data mean. Specify the respective weights.

   Answer:

$$\begin{aligned} \text{posterior mean} &= w\text{prior mean} + (1-w)\text{ data mean} \\ \frac{x+\alpha}{n+\alpha+\beta} &= w\,\frac{\alpha}{\alpha+\beta} + (1-w)\,\frac{x}{n} \end{aligned}$$

   Solve w.r.t. w:

$$\frac{x+\alpha}{n+\alpha+\beta} = \frac{\alpha+\beta}{n+\alpha+\beta}\,\frac{\alpha}{\alpha+\beta} + \frac{n}{n+\alpha+\beta}\,\frac{x}{n}$$

   i.e.

$$w = \frac{\alpha+\beta}{n+\alpha+\beta}.$$

   Here, $w = \frac{5}{121} = 0.041$.

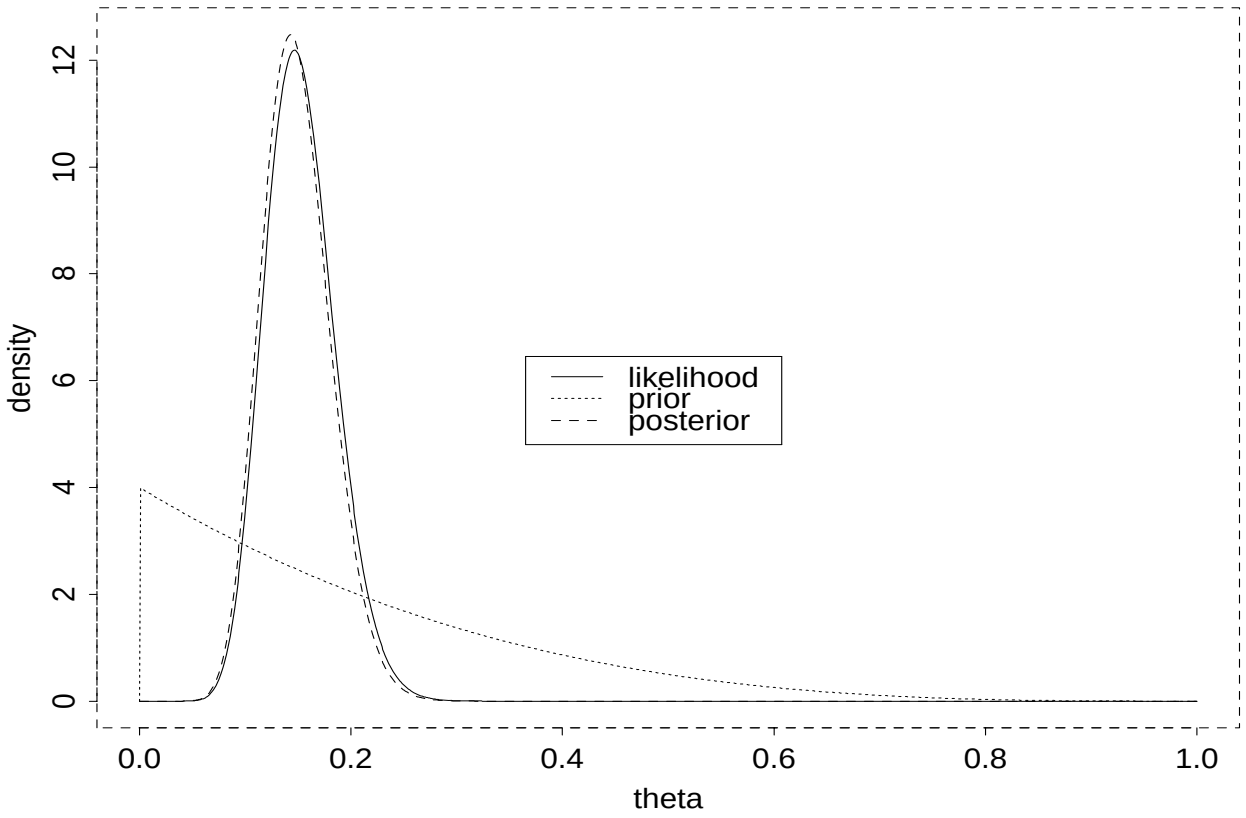

Figure 3: Prior, normalized likelihood, and posterior density of θ.

(e) **[1m]** Plot the prior, posterior and normalized likelihood in one display.

Answer: See Figure 3.

(f) **[1m]** Find the posterior probability that $\theta < 0.1$.

Answer:
$P(\theta < 0.1|x = 17) =$ cdf of $Beta(18, 103)$ at 0.1 $= 0.05309438$ using the Splus function `pbeta(0.1,18,103)`.

(g) **[1m]** Find a central 95% posterior credible interval for θ.

Answer:
Using the Splus functions `qbeta(0.025,18,103)` and `qbeta(0.975,18,103)` gives the 95% central posterior credible interval of [0.09138957, 0.2171069].

(h) **[3m]** Suppose that NIWA plans another study of $n^* = 50$ water samples as above. What is the posterior predictive probability that $x^* = 5$ of these contain Giardia cysts? Derive the formula for general n^*, x^* first.
(Hint: You will need to compute the integral of the kernel of a beta density. Look at the expression for the normalized beta density, and recall that it has to integrate to one.)

Answer:

$$\begin{aligned}
f(x^*|x) &= \int_0^1 \binom{n^*}{x^*} \theta^{x^*}(1-\theta)^{n^*-x^*} f(\theta|x)d\theta \\
&= \int_0^1 \binom{n^*}{x^*} \theta^{x^*}(1-\theta)^{n^*-x^*} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}d\theta \\
&= \binom{n^*}{x^*} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \int_0^1 \theta^{\alpha+x+x^*-1}(1-\theta)^{\beta+n-x+n^*-x^*-1}d\theta \\
&= \binom{n^*}{x^*} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \frac{\Gamma(\alpha+x+x^*)\Gamma(\beta+n-x+n^*-x^*)}{\Gamma(\alpha+\beta+n+n^*)} \\
&= \text{Beta-bin}(n^*, \alpha+x, \beta+n-x)
\end{aligned}$$

For $x^* = 5, \alpha = 1, \beta = 4, n = 116, x = 17$ using the functions `gamma(a,b)` and `beta(a,b)` in R

```
> choose(50,5)*beta(18,103)*beta(23,148)
[1] 0.1108717
```

gives: $f(x^* = 5|\theta, x) = 0.1108717$.

(i) **[2m]** Test the hypothesis

$$H_0 : \theta \geq 0.2 \qquad \text{versus} \qquad H_1 : \theta < 0.2$$

at the 5% significance level in a frequentist manner and interpret of your results.

Answer:
The P-value is the probability to observe 17 out of 116 or less, given that H_0 is true (or given $\theta = 0.2$). Thus,

$$P - value = P(X \leq 17|\theta = 0.2) = pbinom(17, 116, 0.2) = 0.0894762.$$

This does not mean that H_0 is true with probability 0.089, but that assuming that the null-hypothesis is true (i.e. the contamination probability is larger than 0.2) we would observe something as extreme or even more extreme than our observation only with probability of 0.089. As this is larger than the significance level 0.05, we cannot reject the null hypothesis or there is insufficient evidence to reject H_0.

(j) **[2m]** Test the same hypotheses as in part (i) in a Bayesian manner and interpret your results.

Answer:
We calculate the posterior probability of H_0, i.e.

$$P(H_0|x = 17) = P(\theta \geq 0.2|x = 17) = 1 - pbeta(0.2, 18, 103) = 0.06472686.$$

This means that after observing 17 contaminated samples out of a total of 116, the true contamination probability is larger than 0.2 only with a probability 0.065. With probability 0.935, it is larger than 0.2.

7. **[5 marks] Poisson-Exponential**
Suppose we observe a Poisson process with unknown rate θ. The Poisson distribution describes the number of events occurring in a unit time interval. The Exponential distribution describes the waiting time until the next event.

We can obtain information about θ from either

- observing the number of events, or
- observing the waiting time.

Let Y_t denote the random variable describing the number of events occurring in t time units and T the time until the next event.

a) **[2m]** If $Y_t \sim \text{Poisson}(\theta \cdot t)$, show that T has an Exponential(θ) distribution by first calculating the probability that no event occurs in t time units.

Answer:
$Y_t \sim \text{Poisson}(\theta \cdot t)$. Then

$$\begin{aligned} \mathbb{P}(\text{no event in t time units}) &= \mathbb{P}(Y_t = 0) \\ &= \frac{(\theta \cdot t)^0}{0!} \mathrm{e}^{-\theta \cdot t} \\ F_T(t) &= \mathbb{P}(T \le t) = 1 - \mathbb{P}(\text{no events in t time units}) \\ &= 1 - \mathrm{e}^{-\theta \cdot t} \end{aligned}$$

By differentiation

$$f_T(t) = \theta \mathrm{e}^{-\theta \cdot t}, \quad t \ge 0$$

i.e. pdf of Exponential(θ) distribution.

b) **[3m]** Using the same Gamma(α, β) prior for θ in the Poisson and Exponential model, do we end up with the same posterior for θ?

Answer:
Let $T \sim \text{Exponential}(\theta)$ and we observe $T = t$. Using a Gamma(α, β) prior for θ, we obtain (see lectures) a Gamma($\alpha + 1, \beta + t$) posterior distribution for θ.

On the other hand, if $Y_t \sim \text{Poisson}(\theta \cdot t)$ and we observe $Y_t = 1$, then

$$\begin{aligned} f(\theta|y) &\propto f(\theta) f(y|\theta) \\ &\propto \theta^{\alpha-1} \mathrm{e}^{-\beta\theta} (\theta \cdot t)^y \mathrm{e}^{-\theta t} \\ &\propto \theta^{\alpha-1+y} \mathrm{e}^{-\theta(\beta+t)} \end{aligned}$$

i.e. pdf of Gamma($\alpha + y, \beta + t$). If $y = 1$, then the posterior of θ is Gamma($\alpha + 1, \beta + t$).
(For the purpose of finding the posterior distribution of θ it is equivalent to observe either $Y_t = 1$ or $T = t$.)

8. **[6 marks] Binomial Example using WinBugs**
Now make use of WinBugs to answer parts (c), (e) (only plotting the posterior pdf), (f), (g), and (h) of Question 6. Use a burn-in of 1000 iterations and base your answers on 2000 MCMC iterations. Hint: Make use of the `step` function in WinBUGS.

Answer:
WinBUGS code:

```
model
{
theta ~ dbeta(alpha,beta)
x ~ dbin(theta,n)

prob<- 1-step(theta-0.1)  #part (f)

xstar ~ dbin(theta,nstar)  #part (h)
xstar5<-step(xstar-5)-step(xstar-6)
}

list(alpha=1,beta=4, x=17,n=116,nstar=50,xstar=NA) #data
list(theta=.5)   #initial values
```

WinBUGS output:

```
Node statistics
 node  mean  sd  MC error 2.5% median 97.5% start sample
prob 0.0545 0.227 0.004869 0.0 0.0 1.0 1001 2000
theta 0.149 0.03235 8.402E-4 0.09121 0.1461 0.2131 1001 2000
xstar5 0.1065 0.3085 0.006613 0.0 0.0 1.0 1001 2000
```

Thus, the estimated probability that $\theta < 0.1$ is 0.0545 and a 95% posterior credible interval for θ is given by $[0.09121, 0.2131]$.

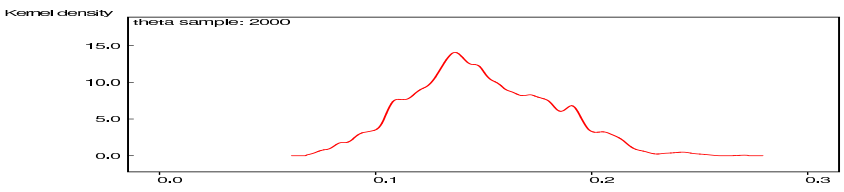

Figure 4: Kernel density estimate of θ.

9. **[30 marks] Dugongs Data**
Carlin and Gelfand (1991) present a nonconjugate Bayesian analysis of the following data set from Ratkowsky (1983):

Dugong (i)	1	2	3	4	$\cdots$	26	27
Age (x_i)	1.0	1.5	1.5	1.5	$\cdots$	29.0	31.5
Length (y_i)	1.80	1.85	1.87	1.77	$\cdots$	2.72	2.57

The data are length (in meters) and age (in years) measurements of 27 captured dugongs (sea cows) of the sirenian species. Carlin and Gelfand (1991) model this data using a nonlinear growth curve with no inflection point and an asymptote as x_i tends to infinity. The model is given by:

$$Y_i = \alpha - \beta \cdot \gamma^{x_i} + \epsilon_i, \quad i = 1, \ldots, n$$

where $\epsilon \sim N(0, \sigma^2)$. In this model, α corresponds to the average length of a fully grown dugong ($x \to \infty$), $\alpha - \beta$ is the length of a dugong at birth ($x = 0$) and γ determines the growth rate: lower values produce an initially steep growth curve while higher values lead to gradual, almost linear growth. The dataset `dugong.txt` can be found on the course webpage.

a) **[8 marks] Simple Linear Regression in WinBUGS**

i) **[1m]** To avoid a nonlinear model, transform the x_i to the log scale and plot y_i versus $log(x_i)$ in R. Does the relationship look reasonably linear?

Answer: R-Code and see Figure 5:

```
x = c( 1.0,  1.5,  1.5,  1.5, 2.5,   4.0,  5.0,  5.0,  7.0,
            8.0,  8.5,  9.0,  9.5, 9.5,  10.0, 12.0, 12.0, 13.0,
           13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5)
Y = c(1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47,
          2.19, 2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43,
          2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57)
z<-log(x)
pdf("dugongslin.pdf")
plot(z,Y,xlab="log(x)",ylab="Y")
dev.off()
```

ii) **[5m]** Using a simple linear regression model to capture the relationship between log(age) and length:

$$Y_i = \beta_0 + \beta_1 \cdot \log(x_i) + \epsilon_i, \quad i = 1, \ldots, n$$

with $\epsilon_i \sim N(0, \sigma^2)$ and standard noninformative priors for β_0, β_1 and σ^2, obtain the posterior distribution of the parameters using WinBUGS. Use a burnin of 1000 iterations and monitor the parameters for a further 5000 iterations. Give posterior summaries of all parameters.
Compare the traceplots, autocorrelations and the cross correlation of (β_0, β_1) to a model with centered covariate.
Answer:

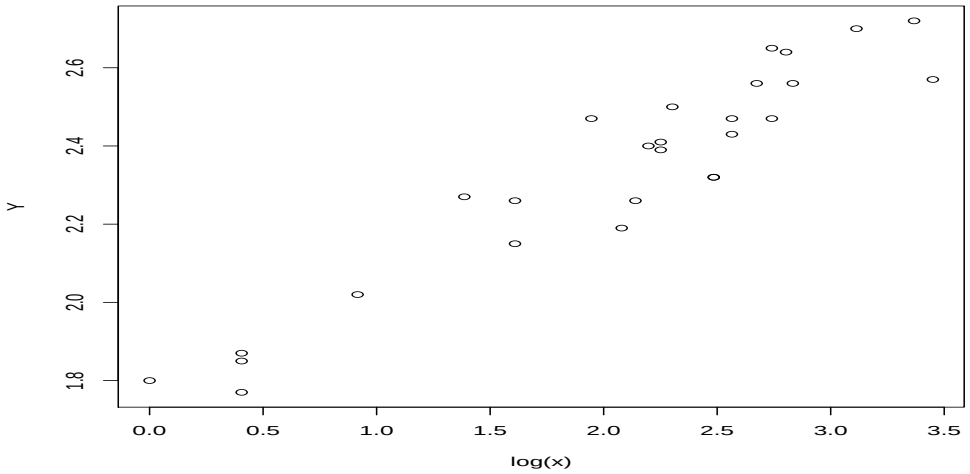

Figure 5: Plot of length versus log-transformed age of dugongs.

```
####################
# Dugongs Linear model
####################

model{
  for( i in 1:n) {
    logage[i] <- log(x[i])
    Y[i] ~ dnorm(mu[i] , tau)
    mu[i] <- beta0+ beta1*logage[i]    # uncentered covariate
#    mu[i] <- beta0+ beta1*(logage[i] - mean(logage[]))    #  centered covariate
    }

 beta0 ~ dnorm(0, 0.001)
 beta1 ~ dnorm(0, 0.001)

  tau ~ dgamma(0.01, 0.01)
  sigma <- 1/sqrt(tau)

#  tau <- 1/(sigma*sigma)
#  sigma ~ dunif(0.01, 100)

mu.new<- beta0 + beta1*log(25)
Y.new ~ dnorm(mu.new,tau)
}  #  end of BUGS code


#Data:
list(x = c( 1.0,  1.5,  1.5,  1.5, 2.5,   4.0,  5.0,  5.0,  7.0,
```

```
            8.0,  8.5,  9.0,  9.5, 9.5,  10.0, 12.0, 12.0, 13.0,
           13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5),
     Y = c(1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47,
           2.19, 2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43,
           2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57), n = 27)

#Data with missing value:
list(x = c( 1.0,  1.5,  1.5,  1.5, 2.5,   4.0,  5.0,  5.0,  7.0,
            8.0,  8.5,  9.0,  9.5, 9.5,  10.0, 12.0, 12.0, 13.0,
           13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5),
     Y = c(1.80, 1.85, NA, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47,
           2.19, 2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43,
           2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57), n = 27)
# Inits:

list( beta0 = 0, beta1 = 1, tau = 1)   #  for gamma prior on tau
list( beta0 = 0, beta1 = 1, sigma = 1)  # for uniform prior on sigma
```

Results:

```
 node  mean  sd  MC error 2.5/% median 97.5/% start sample
beta0 1.76 0.04613 6.333E-4 1.67 1.76 1.85 1001 5000
beta1 0.2778 0.02024 2.9E-4 0.2377 0.2776 0.3185 1001 5000
sigma 0.09743 0.01466 2.267E-4 0.07423 0.09571 0.1313 1001 5000
```

iii) [**1m**] Obtain the posterior predictive distribution of the length of a dugong of age 25.
Answer:

```
 node  mean  sd  MC error 2.5\% median 97.5\% start sample
Y.new 2.655 0.1024 0.001212 2.454 2.652 2.86 1001 5000
```

iv)) [**1m**] Now change the data file to make the observation of y_3 a missing value. Rerun the WinBUGS code, now monitoring y_3 to obtain the predictive distribution.
Answer:

```
 node  mean  sd  MC error 2.5% median 97.5% start sample
Y[3] 1.872 0.1107 0.00165 1.653 1.872 2.089 1001 5000
```

b) **[12 marks] Nonlinear Regression in WinBUGS**

i) **[4m]** Using the untransformed data and nonlinear model with a diffuse Normal prior for α and β, a Uniform(0.5,1) prior for γ, and a Uniform(0.01,100) prior for the standard deviation σ (as suggested by Gelman to bound the prior away from 0 and ∞), obtain the posterior distribution of the parameters with WinBUGS. Use a burnin period of 1000 and a total of 11.000 iterations. Summarize the posterior distribution of all parameters. Obtain the autocorrelation plots and investigate the bivariate correlation of (α, γ) using the `Correlation` tool on the `Inference` menu.

: WinBUGS Code:

```
######################
# Dugongs Nonlinear model
#  See Carlin and Louis (2008), Example 4.3
######################

model
{
for( i in 1 : N ) {
Y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha - beta * pow(gamma,x[i])
}
alpha ~ dnorm(0, 0.001)
beta ~ dnorm(0, 0.001)
gamma ~ dunif(0.5, 1.0)

        tau <- 1/(sigma*sigma)
        sigma ~ dunif(0.01, 100)
}
#Data:
list(x = c( 1.0,  1.5,  1.5,  1.5, 2.5,   4.0,  5.0,  5.0,  7.0,
            8.0,  8.5,  9.0,  9.5, 9.5,  10.0, 12.0, 12.0, 13.0,
           13.0, 14.5, 15.5, 15.5, 16.5, 17.0, 22.5, 29.0, 31.5),
     Y = c(1.80, 1.85, 1.87, 1.77, 2.02, 2.27, 2.15, 2.26, 2.47,
           2.19, 2.26, 2.40, 2.39, 2.41, 2.50, 2.32, 2.32, 2.43,
           2.47, 2.56, 2.65, 2.47, 2.64, 2.56, 2.70, 2.72, 2.57), N = 27)
# Inits:
list(alpha = 1, beta = 1, sigma = 1, gamma = 0.9)
list(alpha = 10, beta = 10, sigma = 10, gamma = 0.7)
list(alpha = 100, beta = 100, sigma = 100, gamma = 0.5)
```

Posterior summary:

```
 node  mean  sd  MC error 2.5% median 97.5% start sample
alpha 2.656 0.07568 0.003196 2.529 2.647 2.829 1001 10000
beta 0.9744 0.07842 0.001198 0.8232 0.9736 1.136 1001 10000
gamma 0.8634 0.03357 0.001443 0.7864 0.8669 0.919 1001 10000
```

```
sigma 0.1008 0.01555 2.204E-4 0.07616 0.09896 0.1372 1001 10000
```

ii) **[2m]** In R, plot the data and the fitted growth curve using the posterior means as point estimates/fitted values of the parameters.

Answer: See Figure 6

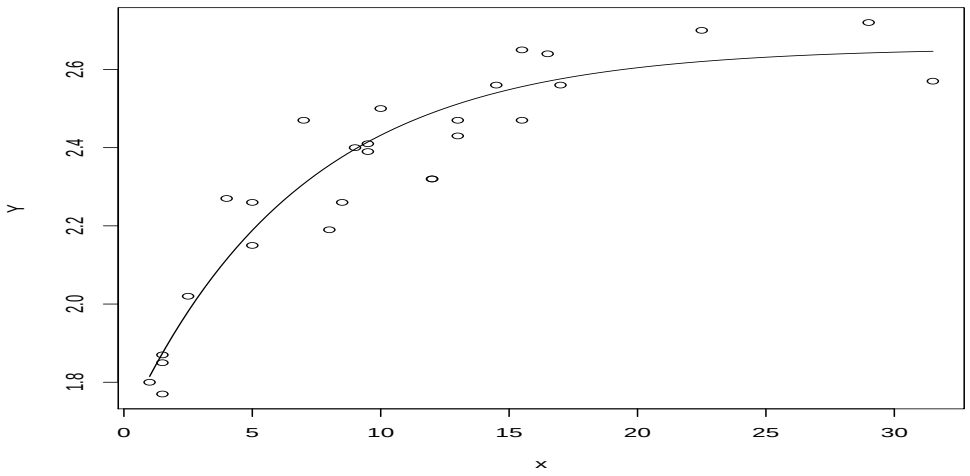

Figure 6: Plot of length versus age of dugongs with fitted growth curve.

iii) **[5m]** Perform all model checking criteria in WinBUGS as described in Chapter 4.3.

Answer: WinBUGS code

```
######################
# Dugongs Nonlinear model: Model checking
######################

model
{
PI <- 3.14159
for( i in 1 : N ) {
Y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha - beta * pow(gamma,x[i])

#residuals
r[i]  <- Y[i]-mu[i]
sr[i]<- (Y[i]-mu[i])*sqrt(tau)
m3[i] <- pow(sr[i],3)
m4[i] <- pow(sr[i],4)


y.rep[i] ~dnorm(mu[i],tau)
p.smaller[i]<-step(Y[i]-y.rep[i])
```

```
like[i] <- sqrt(tau)/(2*PI)*exp(-0.5*pow(sr[i],2))
p.inv[i] <- 1/like[i]

#residuals of replicates
resid.rep[i] <- y.rep[i]-mu[i]
sresid.rep[i]<- resid.rep[i]*sqrt(tau)
m3.rep[i] <- pow(sresid.rep[i],3)
m4.rep[i] <- pow(sresid.rep[i],4)
}

alpha ~ dnorm(0, 0.001)
beta ~ dnorm(0, 0.001)
gamma ~ dunif(0.5, 1.0)

        tau <- 1/(sigma*sigma)
        sigma ~ dunif(0.01, 100)

#bayesian p-value:
skew.obs <- sum(m3[])/N
         skew.rep <- sum(m3.rep[])/N
         p.skew   <-  step(skew.rep-skew.obs)

         kurtosis.obs <- sum(m4[])/N
         kurtosis.rep <- sum(m4.rep[])/N
         p.kurtosis    <- step(kurtosis.rep-kurtosis.obs)
}
```

c) **[10 marks] Logistic Regression in WinBUGS**
Now consider a binary version of the dugong data,

$$Z_i = \begin{cases} 1 & \text{if } Y_i > 2.4 \quad \text{i.e. if the dugong is fully grown} \\ 0 & \text{otherwise.} \end{cases}$$

A logistic model for $\pi = \mathbb{P}(Z_i = 1)$ is then

$$logit(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 \log(x_i).$$

i) Fit this GLM with WinBUGS using the standard noninformative priors for β_0 and β_1. Also fit models using the probit (use the `phi` function instead of the less stable `probit` function) and the complementary-log-log link functions.

ii) Compare π_i boxplots (induced by different link functions and the β_0 and β_1 posteriors) using the `Comparison` tool.

iii) Compare the fit for different link functions using the deviance information criterion.
Answer: code in `dugongsBin_BUGS.txt`

10. [**12 marks**] **CD4 counts**
The data for this problem are in the plain text file "CD411b.txt" on the course webpage. It consists of transformed CD4 counts on 26 patients that were taken at weeks 0,2,8,12 and 24 after they began treatment in an AIDS clinical trial called ACTG 116B. The 4-th root transformation has been applied to stabilize variance. This is often done with CD4-count data.

a) **[4m]** Use the "Rats" example in the WinBUGS example volume 1 as a template to fit a hierarchical linear model to these data. (You may need to choose different starting values nd possibly different priors than those that worked for the "Rats" data.)

b) **[3m]** Assess convergence using `coda`.

c) **[5m]** Give estimated posterior means and 95% credible intervals for the following quantities:

i) The estimated population intercept of 4th root CD4 for all patients at the beginning of the study (i.e. at week 0). (If you center the covariate, think about how to do this.)

ii) The estimated population slope of 4th root CD4 values around the regression line.

iii) The standard deviation of 4th root CD4 values around the regression line.

iv) The standard deviation that captures between-patient variability in slopes.

v) The individual intercept and slope for patient 12.

Answer: WinBUGS code:

```
model
{
for( i in 1 : N ) {
for( j in 1 : T ) {
Y[i , j] ~ dnorm(mu[i , j],tau.c)
mu[i , j] <- alpha[i] + beta[i] * (x[j] - xbar)
}
alpha[i] ~ dnorm(alpha.c,tau.alpha)
beta[i] ~ dnorm(beta.c,tau.beta)
}
 #tau.c ~ dgamma(0.1,0.1)
 #  sigma <- 1 /sqrt(tau.c)
 sigma ~ dunif(0.01,100)
     tau.c<-1/sigma/sigma
alpha.c ~ dnorm(0.0,1.0E-6)
beta.c ~ dnorm(0.0,1.0E-6)
# Choice of prior of random effects variances
# Prior 1: uniform on SD
sigma.alpha~ dunif(0.01,100)
sigma.beta~ dunif(0.01,100)
tau.alpha<-1/(sigma.alpha*sigma.alpha)
tau.beta<-1/(sigma.beta*sigma.beta)
alpha0 <- alpha.c - xbar * beta.c
}
```

WinBUGS script:

```
display('log')
check('PATHNAME/CD4_BUGS.odc')
data('PATHNAME/CD4116b.txt')
compile(2)
inits(1,'PATHNAME/CD4_inits1.txt')
gen.inits()
inits(2,'PATHNAME/CD4_inits2.txt')
gen.inits()
update(1000)
set(alpha0)
set(alpha[12])
set(beta[12])
set(sigma)
set(alpha.c)
set(sigma.alpha)
set(beta.c)
set(sigma.beta)
update(5000)
stats(*)
history(*)
coda(*,'PATHNAME/CD4_output')
save('PATHNAME/CD4_log')
```

WinBUGS output:

```
 node  mean  sd  MC error 2.5% median 97.5% start sample
alpha[12] 3.705 0.2006 0.001985 3.309 3.707 4.096 1001 10000
alpha.c 2.515 0.1464 0.001603 2.228 2.514 2.798 1001 10000
alpha0 3.089 0.1603 0.002196 2.77 3.088 3.405 1001 10000
beta[12] -0.06551 0.01503 2.247E-4 -0.09679 -0.06492 -0.03755 1001 10000
beta.c -0.05578 0.006339 1.517E-4 -0.06826 -0.05586 -0.04295 1001 10000
sigma 0.5061 0.03686 5.032E-4 0.4406 0.504 0.5843 1001 10000
sigma.alpha 0.7006 0.1194 0.001612 0.507 0.6866 0.9718 1001 10000
sigma.beta 0.01693 0.005312 1.603E-4 0.01029 0.01582 0.02981 1001 10000
```

11. **[6 marks] Epilepsy data**

A neurologist has four patients with epilepsy, all of whom are taking the same drug to control their seizures. They have been on the drug long enough that their pattern of seizures has stabilized.

The neurologist asks each patient to keep track of the number of seizures he or she has in each of three months. Let x_{ij} represent the number of seizures that patient i has in month j, where i goes from 1 to 4 and j goes from 1 to 3.

The neurologist believes that each patient has a true mean number of seizures θ_i per month, and that the number of seizures that each patient has in a given month is a draw from a Poisson distribution with parameter θ_i. That is

$$X_{ij}|\theta_i \sim \text{Poisson}(\theta_i), \quad i = 1, \ldots, 4, j = 1, \ldots, 3,$$

and the X_{ij}s are considered conditionally independent given θ_i.

Furthermore, she believes that, although the θ_is are different for different patients, they are similar, and that that similarity can be expressed by saying that all the θ_is are like draws from a common Gamma distribution. That is

$$\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta), \quad i = 1, \ldots, 4.$$

Finally, she believes that the following prior densities are appropriate for the parameters α and β:

$$\begin{aligned} \alpha &\sim \text{Exponential}(0.1) \\ \beta &\sim \text{Gamma}(0.1, 1) \end{aligned}$$

The neurologist carries out a Bayesian analysis to learn about each individual patient's mean parameter θ_i and about the density from which the θ_is are drawn. She uses WinBUGS to run three MCMC chains using different sets of initial values.

a) **[3m]** Draw a DAG of the model and let WinBUGs generate the code.

Answer: see Figure 9.

b) **[1m]** Which quantities are treated as exchangeable in this model?

Answer: x_{i1}, x_{i2}, x_{i3} for all i are exchangeable and $\theta_1, \theta_2, \theta_3, \theta_4$ are exchangeable.

c) **[1m]** Download the data file `epil.txt` from the course webpage and run the model for 5000 iterations without burn-in. How many iterations does the Gelman-Rubin plot for α suggest should be discarded as burn-in?

Answer: There are many possible good answers. We want the red line to get close to 1 and stay there. This happens at about 100 iterations. We want the green and blue line to come together and stabilize. This happens at about after 400 iterations.

d) **[1m]** What is the 95% credible interval for θ_2?

Answer: [1.442,4.871]

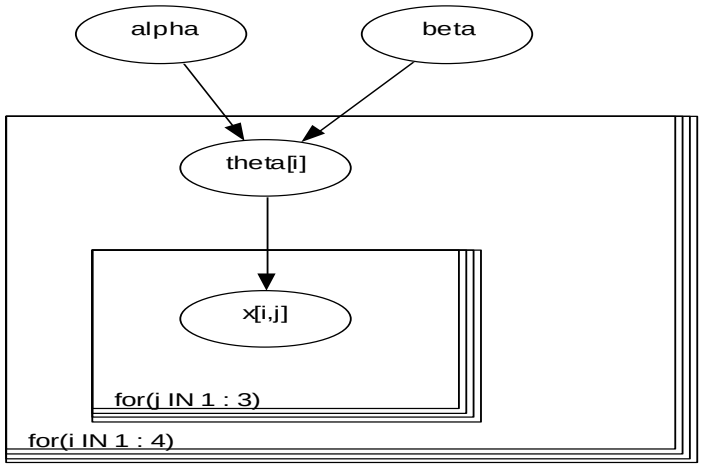

Figure 7: Directed acyclic graph of Epilepsy Model.