

# MAS8303 Modern Bayesian Inference

## Part 2

### TEST with SOLUTION

Semester 1, 2011-2

*Time allowed: ONE HOUR.*

*There are THREE questions. Answer all questions.*

*This is an "OPEN BOOK" test. You may use books, lecture notes etc. but you may not consult anyone other than the invigilator during the test. You may use the computer and a calculator. Statistical tables are not provided but you are allowed to use your own or to use the computer instead.*

*Write your answers in the answer book provided. If you wish to send me any graphs or other files then please email them to me at the following address.*

`malcolm.farrow@newcastle.ac.uk`

The following data are taken from Feigl and Zelen (1965). They refer to 33 leukaemia patients, classified as either "AG positive" or "AG negative". The initial white blood cell counts and the survival times in weeks are given. We are interested in what we can say about how long a patient is likely to survive, given the white blood cell count (WBC) and the AG group.

AG positive		AG negative	
WBC	Survival Time	WBC	Survival Time
2300	65	4400	56
750	156	3000	65
4300	100	4000	17
2600	134	1500	7
6000	16	9000	16
10500	108	5300	22
10000	121	10000	3
17000	4	19000	4
5400	39	27000	2
7000	143	28000	3
9400	56	31000	8
32000	26	26000	4
35000	22	21000	3
100000	1	79000	30
100000	1	100000	4
52000	5	100000	43
100000	65		

Let  $T_i$  be the survival time for patient  $i$ , let  $x_i$  be the natural logarithm of the corresponding white blood cell count and let patient  $i$  be in AG group  $g_i$  where  $g_i = 1$  for AG positive and  $g_i = 2$  for AG negative. Then, given  $x_i$  and  $g_i$  and the model parameters, we suppose that the distribution of  $T_i$  is Weibull with probability density function

$$f_i(t) = \alpha \lambda_i t^{\alpha-1} \exp(-\lambda_i t^\alpha)$$

where

$$\lambda_i = \exp\{\beta_0 + \beta_1(x_i - 10) + \beta_2(2g_i - 3)\}.$$

The cumulative distribution function is

$$F_i(t) = 1 - \exp(-\lambda_i t^\alpha).$$

We give independent prior distributions to the three model parameters as follows.

$$\begin{aligned} \alpha &\sim \text{Ga}(1.2, 0.2), \\ \beta_0 &\sim N(-2, 10), \\ \beta_1 &\sim N(0, 4), \\ \beta_2 &\sim N(0, 4). \end{aligned}$$

1. Write down, in standard mathematical notation, the meaning of the following BUGS expressions.

(a) `z<-log(2)/lambda.p`  
`gamma<-pow(z,1/alpha)`

**SOLUTION :**

$$\begin{aligned} z &= \frac{\ln(2)}{\lambda_p} \\ \gamma &= z^{1/\alpha} \end{aligned}$$

(b) `t[i]~dweib(alpha,lambda[i])`

**SOLUTION :**

$$t_i \sim \text{Weibull}(\alpha, \lambda_i)$$

(4 marks)

2. Use BRugs to evaluate the posterior distribution. *Read the rest of this question to see what you need to do before you run the program.*

- (a) Write down your BRugs model specification.

**SOLUTION :**

```
model test12

{for (i in 1:33)
  {t[i]~dweib(alpha,lambda[i])
   lambda[i]<-exp(beta0+beta[1]*(log(wbc[i])-10)+beta[2]*(2*ag[i]-3))
  }

alpha~dgamma(1.2,0.2)
beta0~dnorm(-2,0.1)
beta[1]~dnorm(0,0.25)
beta[2]~dnorm(0,0.25)

lambdastar<-exp(beta0+beta[1]*(log(1000)-10)-beta[2])
z<-log(2)/lambdastar
gamma<-pow(z,1/alpha)

}
```

(b) Write down the commands which you use.

**SOLUTION** : See the following session log.

```
> library(BRugs)
Loading required package: coda
Loading required package: lattice
Welcome to BRugs running on OpenBUGS version 3.0.3
> modelCheck("test12bug1.txt")
model is syntactically correct
> modelData("leuksurvdata.txt")
data loaded
> modelCompile(2)
model compiled
> modelGenInits()
initial values generated, model initialized
> samplesSet(c("alpha","beta0","beta"))
monitor set for variable 'alpha'
monitor set for variable 'beta0'
monitor set for variable 'beta'
> modelUpdate(2000)
2000 updates took 1 s
> samplesHistory("alpha")
Waiting to confirm page change...
> samplesHistory("beta0")
Waiting to confirm page change...
> samplesHistory("beta")
Waiting to confirm page change...
> modelCheck("test12bug1.txt")
model is syntactically correct
> modelData("leuksurvdata.txt")
data loaded
> modelCompile(2)
model compiled
> modelGenInits()
initial values generated, model initialized
> modelUpdate(10000)
10000 updates took 6 s
> samplesSet(c("alpha","beta0","beta"))
monitor set for variable 'alpha'
monitor set for variable 'beta0'
monitor set for variable 'beta'
> samplesSet("gamma")
monitor set for variable 'gamma'
> modelUpdate(20000)
20000 updates took 13 s
> samplesHistory("alpha")
Waiting to confirm page change...
> samplesHistory("beta0")
Waiting to confirm page change...
> samplesHistory("beta")
Waiting to confirm page change...
> samplesHistory("gamma")
Waiting to confirm page change...
```

```

> samplesStats(c("alpha","beta0","beta","gamma"))
      mean      sd MC_error val2.5pc  median val97.5pc start sample
alpha   0.9723  0.1339 0.005852  0.71950  0.9707    1.2380 10001 40000
beta0  -3.2340  0.5416 0.025020 -4.32400 -3.2440   -2.2290 10001 40000
beta[1] 0.3012  0.1287 0.001555  0.04768  0.3024    0.5516 10001 40000
beta[2] 0.4755  0.1948 0.003648  0.09707  0.4720    0.8582 10001 40000
gamma   87.2800 37.5100 0.406500 36.86000 80.1900  180.1000 10001 40000

```

(c) Find the posterior mean of  $\beta_0$ .

**SOLUTION** : The posterior mean of  $\beta_0$  is  $-3.2340$ .

(d) Find a 95% posterior interval for  $\alpha$ .

**SOLUTION** : A 95% posterior interval for  $\alpha$  is  $0.71950 < \alpha < 1.2380$ .

(e) Let  $\gamma(\alpha, \beta_0, \beta_1, \beta_2)$  be the median lifetime, given the parameters, for an AG positive patient with white blood cell count 1000. Show that

$$\gamma = \left( \frac{\ln 2}{\lambda_p} \right)^{1/\alpha}$$

where  $\lambda_p = \exp[\beta_0 + \beta_1 \{\ln(1000) - 10\} - \beta_2]$  and find the posterior mean of  $\gamma$ .

**SOLUTION** : We have  $x_i = \ln 1000$  and  $g_i = 1$ . So

$$\lambda_i = \exp[\beta_0 + \beta_1 \{\ln(1000) - 10\} - \beta_2] = \lambda_p.$$

The distribution function, evaluated at  $t = \gamma$ , is

$$F(\gamma) = 1 - \exp(-\lambda_p \gamma^\alpha) = 1/2.$$

Hence  $\exp(-\lambda_p \gamma^\alpha) = 1/2$  which leads to  $\lambda_p \gamma^\alpha = \ln 2$  and

$$\gamma = \left( \frac{\ln 2}{\lambda_p} \right)^{1/\alpha}.$$

The posterior mean of  $\gamma$  is 87.28 weeks.

(f) Comment on the behaviour of the MCMC sampler (e.g. convergence, mixing).

**SOLUTION** : Convergence and mixing are rather poor. A burn-in of at least 1000 iterations, preferably more, is required. See the trace plots in Figures 1, 2 and 3.

(10 marks)

**Data.** A ready-to-use data file is available on the module Web page at the following address.

<http://www.mas.ncl.ac.uk/~nmf16/teaching/mas8303/>

It is listed under “Data” and “Leukaemia survival data” as `leuksurvdata.txt`. The contents are as follows. (Recall that there are 33 observations.)

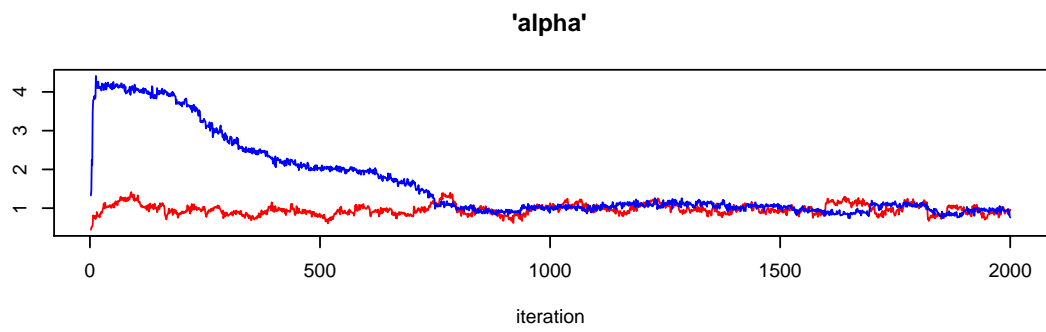


Figure 1: Trace plot for  $\alpha$ .

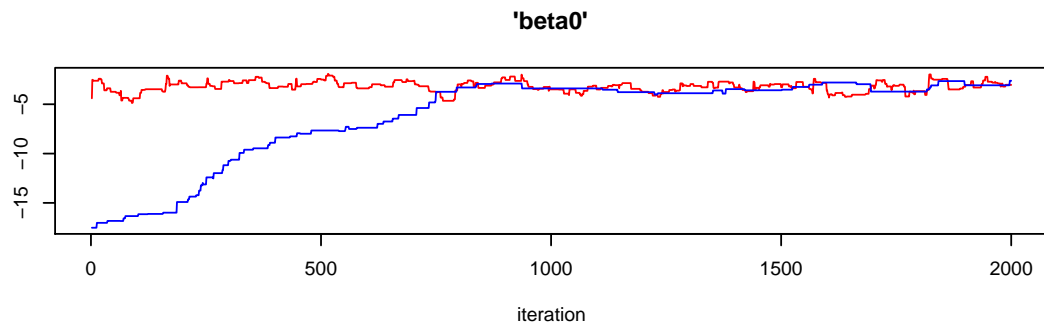


Figure 2: Trace plot for  $\beta_0$ .

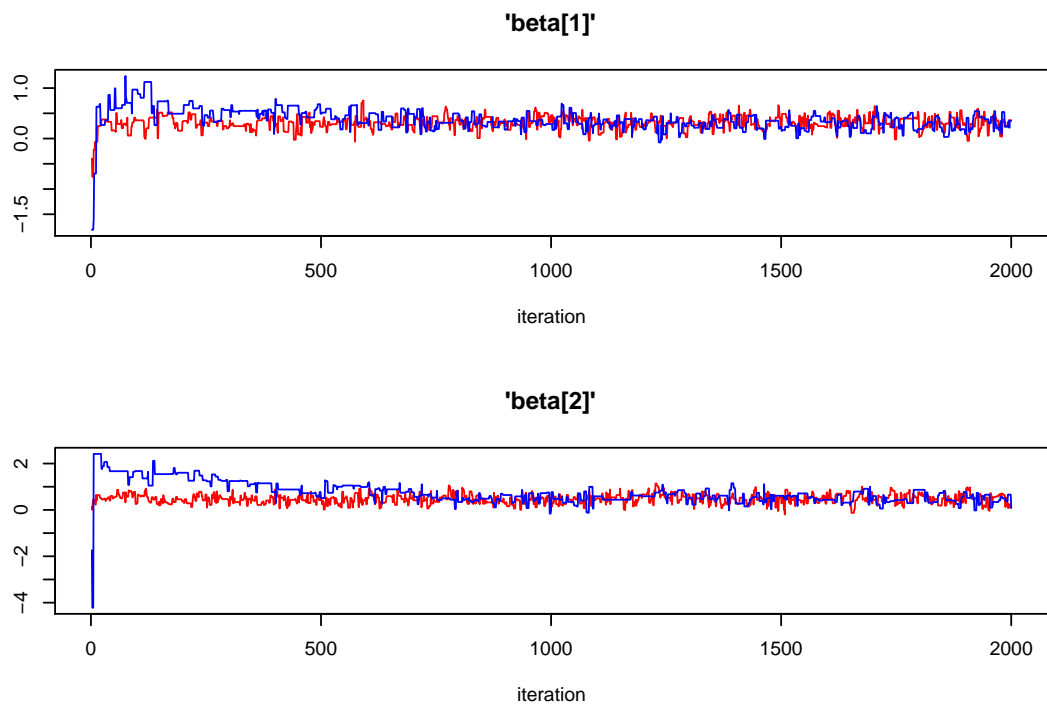


Figure 3: Trace plots for  $\beta$ .

wbc[]	t[]	ag[]
2300	65	1
750	156	1
4300	100	1
2600	134	1
6000	16	1
10500	108	1
10000	121	1
17000	4	1
5400	39	1
7000	143	1
9400	56	1
32000	26	1
35000	22	1
100000	1	1
100000	1	1
52000	5	1
100000	65	1
4400	56	2
3000	65	2
4000	17	2
1500	7	2
9000	16	2
15300	22	2
10000	3	2
19000	4	2
27000	2	2
28000	3	2
31000	8	2
26000	4	2
21000	3	2
79000	30	2
100000	4	2
100000	43	2

END

3. We wish to predict the survival time for an AG positive patient with white blood cell count 1000.

**SOLUTION** : We add an extra “dummy” patient to the data file, with the required co-variates and a missing, *ie* NA , survival time. Then we modify the model specification to have 34 patients and use the commands in the following session log.

```

> modelCheck("test12bug2.txt")
model is syntactically correct
> modelData("leuksurvdata2.txt")
data loaded
> modelCompile(2)
model compiled
> modelGenInits()
initial values generated, model initialized
> modelUpdate(10000)
10000 updates took 6 s
> samplesSet("t")
monitor set for variable 't'
> modelUpdate(20000)

```



```
20000 updates took 13 s
> samplesHistory("t")
Waiting to confirm page change...
> samplesStats("t")
      mean    sd MC_error val2.5pc median val97.5pc start sample
t[34] 132.4 172.3   1.022   2.139   79.7    575.4 10001 40000
```

(a) Find posterior predictive median lifetime for this patient.

**SOLUTION** : The posterior predictive median lifetime for this patient is 79.7 weeks.

(b) Find a 95% posterior predictive interval for the lifetime of this patient.

**SOLUTION** : A 95% posterior predictive interval for the lifetime of this patient, in weeks, is  $2.139 < t < 574.4$ .

(6 marks)

## Reference

Feigl, P. and Zelen, M. (1965) Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826-38.