

# **BAYESIAN STATISTIC COURSE**

**EDINBURGH, SEPTEMBER 4-7<sup>TH</sup> 2007**

**COMPUTER LAB NOTES**

**AGUSTÍN BLASCO  
PROGRAMS BY WAGDY MEKKAWY**

## **AN INTRODUCTION TO BAYESIAN ANALYSIS AND MCMC**

**Agustín Blasco**

Departamento de Ciencia Animal.  
Universidad Politécnica de Valencia  
P.O. Box 22012. Valencia 46071. Spain  
[ablasco@dca.upv.es](mailto:ablasco@dca.upv.es)  
[www.dcam.upv.es/dcia/ablasco](http://www.dcam.upv.es/dcia/ablasco)

# 1. “FIXED” EFFECTS MODEL

## EXAMPLE 1: Litter size data

EXAMPLE 1: A divergent selection experiment on litter size has been performed and we have data of litter size from the High and Low lines. We want to compare both lines at constant ovulation rate; i.e., we want the response to selection not dependent on possible differences in ovulation rate between lines. Data have been measured in two seasons.

### MODEL

$$y_{ikm} = \mu + \text{Line}_i + \text{Season}_k + b \cdot \text{OvulationRate}_{ikm} + \epsilon_{ikm}$$

### TRAITS

#### Litter size

#### “FIXED” EFFECTS

##### Line

Line 1 selected for high litter size

Line 2 selected for low litter size

##### Season

Season 1 = Autumn

Season 2 = Summer

### COVARIATE

#### Ovulation rate

N. data = 84

### DATA FILE

It is a text file called **Data.txt**

Line	Season	OvulationRate	LitterSize
2	2	15	1
2	1	11	5

2	1	12	10
2	1	18	8
1	1	14	8
1	1	11	8
2	1	12	4
2	2	14	9
1	2	13	5

PROCEDURE

### PROGRAM Fixed

INPUT

#### Data.txt

It is the data file, in txt format with tabs, with no head in the first line:

2	2	15	1
2	1	11	5
2	1	12	10
2	1	18	8
1	1	14	8
1	1	11	8
2	1	12	4
2	2	14	9
1	2	13	5

#### Files.txt

It is a file needed by the program. Do not modify it.

PROGRAM

#### Fixed

Enter the total number of traits

1

Enter the total number of data

84

Enter the number of 'fixed' effects

2

Enter the number of levels of 'fixed' effect 1

2

Enter the number of levels of 'fixed' effect 2

2

Enter the number of covariates (0.1.2.....)

1

Would you like to introduce the random seed (1=Yes. 2=No)

1

Enter a random seed

1234

Do you want to establish the MCMC characteristics? (1=Yes. 2=No)

1

Enter chain length

60000

Enter burn-in

10000

Enter lag between samples

10

OUTPUT

### DIC.txt

It is a file with the DIC in the first row and some more information.

### Results1.txt

It is a file for the 1st trait (litter size) with:

ITER: number of iteration

VE: chain for the variance of the error

RSD: chain for the Residual standard deviation (Square root of VE)

MEAN: chain for the general Mean of the model

ef1lev1: chain for the level 1 of the "fixed" effect 1

ef1lev2: chain for the level 2 of the "fixed" effect 1

ef1lev3: chain for the level 3 of the "fixed" effect 1

ef2lev1: chain for the level 1 of the "fixed" effect 2

ef2lev2: chain for the level 2 of the "fixed" effect 2

cov1: chain for the covariate

### PROGRAM PostGibbs

INPUT

### Chain.txt

It is a file with two columns, the first one is the number of iteration and the second the MCMC chain to be analyzed. If the MCMC result comes from the BUGS program, we have several chains concatenated; for example, the first chain goes from iteration 1 to 5000, the second chain from iteration 5001 to 10000 and so on.

We will create a file taking ITER and a new column with the contrast [ef1lev1 – ef1lev2] from Results1.txt. This contrast compares line 1 and line 2 for litter size corrected for ovulation rate.

1	0.0193
2	0.015
3	-0.0077
4	-0.0043
5	0.0133
6	0.0068
7	0.0069
8	0.0032

### Figures for inference

We have to decide which the relevant value for the trait is. For Litter size we decide  $k=1$  using economical arguments. This permits to calculate the probability of relevance  $P>k$  or  $P<-k$ . This also produces the interval  $[-k,+k]$  for the probability of similarity. When  $k=0$  we calculate the probability of the contrast being higher or lower than 0.

### PROGRAM

#### PostGibbs

Enter a chain from BUGS

First iteration

1

Last iteration

5000

Enter k for  $P<k$  or  $P>k$

1

Enter the probability for the HPD interval

0.95

Do you want to calculate the probability of an interval  $[a,b]$ ? (Y=1, N=0)

0 (we decide not to calculate the probability of similarity, since we have seen in the screen that they differ in a relevant quantity)

### OUTPUT

#### Features.txt

Mean = 2.0369

Median = 2.0666

the distribution looks symmetrical, since the median and mean are close.

$(P> 1.0000) = 0.96$

$(P< 1.0000) = 0.04$

the lines have a relevant difference with a high probability

HPD 0.95 = [ 0.9920 3.1776 ]

k for the interval  $[k,+inf)$  with probability 95% = 1.0760

(the other interval  $(-inf,k]$  is irrelevant for inference in this case).

we state that the lines differ at least in 1.07 kits (not at least in 0.9920 kits, as the HPD interval would suggest).

MCse = 0.00779

Geweke Z-Score = 0.0419

<b>EXAMPLE 2: Carcass traits data</b>
---------------------------------------

EXAMPLE 2: We want to compare carcass traits of three breeds of rabbits, two of them selected for litter size and one of the for growth rate. We want to do this at constant liveweight.

### MODEL

$$y_{ikm} = \mu + \text{Line}_i + \text{Sex}_k + b \cdot \text{LiveWeight}_{ikm} + e_{ikm}$$

### TRAITS

**Carcass Weight**  
**Perirenal fat (%)**  
**pH**

### “FIXED” EFFECTS

#### Line

Line A selected for litter size = 1  
 Line V selected for litter size = 2  
 Line R selected for growth rate = 3

#### Sex

Males = 1  
 Females = 2

### COVARIATE

#### Liveweight

N. data: 120

### DATA FILE

Line	Sex	LiveWeight	CarcassWeight	Fat%	pH
2	1	2150	1180	2.20	5.53
2	1	2130	1185	2.05	5.55
2	2	2090	1148	1.86	5.46
2	2	1860	1081	1.72	5.43
1	1	1610	879	1.31	5.48

1	1	2320	1250	1.71	5.43
3	2	1900	1036	1.70	5.43
3	2	1940	1061	1.47	5.52

## PROCEDURE

### PROGRAM Fixed

#### INPUT

##### Data.txt

It is the data file, in txt format with tabs, with no head in the first line:

2	1	2150	1180	2.20	5.53
2	1	2130	1185	2.05	5.55
2	2	2090	1148	1.86	5.46
2	2	1860	1081	1.72	5.43
1	1	1610	879	1.31	5.48
1	1	2320	1250	1.71	5.43
3	2	1900	1036	1.70	5.43
3	2	1940	1061	1.47	5.52

##### Files.txt

It is a file needed by the program. Do not modify it.

## PROGRAM

### Fixed

Enter the total number of traits

3

Enter the total number of data

120

Enter the number of 'fixed' effects

2

Enter the number of levels of 'fixed' effect 1

3

Enter the number of levels of 'fixed' effect 2

2

Enter the number of covariates (0.1.2.....)

1

Would you like to introduce the random seed (1=Yes. 2=No)

1

Enter a random seed

1234

Do you want to establish the MCMC characteristics? (1=Yes. 2=No)

1

Enter chain length

60000

Enter burn-in  
10000  
Enter lag between samples  
10

## OUTPUT

### DIC.txt

It is a file with the DIC in the first row and some information more.

### Results1.txt

It is a file for the 1st trait (carcass weight) with:

ITER: number of iteration

VE: chain for the variance of the error

RSD: chain for the Residual standard deviation (Square root of VE)

MEAN: chain for the general Mean of the model

ef1lev1: chain for the level 1 of the "fixed" effect 1

ef1 lev2: chain for the level 2 of the "fixed" effect 1

ef1 lev3: chain for the level 3 of the "fixed" effect 1

ef2 lev1: chain for the level 1 of the "fixed" effect 2

ef2 lev2: chain for the level 2 of the "fixed" effect 2

cov1: chain for the covariate

### Results2.txt

Same file for trait 2 (Perirenal fat %)

### Results3.txt

Same file for trait 3 (muscular pH of l. dorsi)

## PROGRAM PostGibbs

### INPUT

#### Chain.txt

It is a file with two columns, the first one is the number of iteration and the second the MCMC chain to be analyzed. If the MCMC result comes from the BUGS program, we have several chains concatenated; for example, the first chain goes from iteration 1 to 5000, the second chain from iteration 5001 to 10000 and so on.

We will create a file taking ITER and a new column with the contrast [ef1lev1 – ef1lev2] from Results3.txt. This contrast compares breed A and breed V for muscular pH of l. dorsi

1	0.0193
2	0.015
3	-0.0077
4	-0.0043

5	0.0133
6	0.0068
7	0.0069
8	0.0032
9	0.0294

### Figures for inference

We have to decide which the relevant value for the trait is. For pH we decide  $k=0.10$ , corresponding to 1/2 of the SD of the trait. This permits to calculate the probability of relevance  $P>k$  or  $P<-k$ . This also produces the interval  $[-k,+k]$  for the probability of similarity. When  $k=0$  we calculate the probability of the contrast being higher or lower than 0.

### PROGRAM

#### PostGibbs

Enter a chain from BUGS

First iteration

1

Last iteration

5000

Enter k for  $P<k$  or  $P>k$

0

Enter the probability for the HPD interval

0.95

Do you want to calculate the probability of an interval  $[a,b]$ ? (Y=1, N=0)

1

Give the value of a

-0.10

Give the value of b

0.10

### OUTPUT

#### Features.txt

Mean = 0.017

Median= 0.018

The differences in pH between breeds is very small, and the distribution looks symmetrical, since the median are mean are close.

$P>0 = 0.79$

We have some evidence about that line A might have a higher pH than line V. Running PostGibbs again and calculating the probability of this difference to be relevant ( $k=0.05$ ) gives

$P>0.05 = 0.05$

This means that even if pH of breed A is higher than pH of breed V, the probability of this difference being relevant is very low.

HPD 0.95 = [ -0.0217 0.0563 ]

Notice that the interval includes 0, but we still have some evidence about that line A might have a higher pH than line V. Sometimes 0 is included in the HPD95% and at the same time  $P>0$  is higher than 95%.

Probability of interval [ -0.05 , 0.05] = 0.94

This means that we can state that both breeds have the same pH, in practical or biological terms, since the probability of the difference between breeds being lower than the limits of what we consider to be relevant is high (94%). Notice that we can state that pH of line A is higher than pH of line V and at the same time we can state that this is irrelevant, and both pH are the same for the decisions we have to take from our experiment.

The other features are irrelevant for inference in this case.

## 2. "MIXED" MODEL

EXAMPLE: A divergent selection experiment on litter size has been performed and we have data of neonatal survival (number of born alive / litter size) from several parities of does of the High and Low lines. We want to compare both lines. Former analyses suggest that we do not have a parity effect, but we can still have a "random" effect of doe which includes genetics and environmental permanent effects.

MODEL

$$y_{ikm} = \mu + \text{Line}_i + \text{Doe}_{ik} + e_{ikm}$$

TRAITS

**Neonatal survival**

"FIXED" EFFECTS

**Line**

Line 1 selected for high litter size

Line 2 selected for low litter size

"RANDOM" EFFECTS

**Doe**

N. data = 144

N. animals = 73

DATA FILE

Doe	Line	Survival
1	1	0.85
1	1	0.20
2	1	1
2	1	1
3	1	1
3	1	1
4	1	0.50
4	1	0.62

5	2	0.83
5	2	1
6	2	1
6	2	0.71

## PROCEDURE

### PROGRAM Mixed

#### INPUT

##### Data.txt

It is the data file, in txt format with tabs, with no head in the first line:

1	1	0.85
1	1	0.20
2	1	1
2	1	1
3	1	1
3	1	1
4	1	0.50
4	1	0.62
5	2	0.83
5	2	1
6	2	1

##### Files.txt

It is a file needed by the program. Do not modify it.

## PROGRAM

### Mixed

Enter the total number of data

144

Enter the total number of animals with data

73

Enter the number of 'fixed' effects

1

Enter the number of levels of 'fixed' effect 1

2

Enter the number of covariates (0.1.2.....)

0

Would you like to introduce the random seed (1=Yes. 2=No)

1

Enter a random seed

4321

Do you want to establish the MCMC characteristics? (1=Yes. 2=No)

1

Enter chain length  
60000  
Enter burn-in  
10000  
Enter lag between samples  
10

## OUTPUT

### Results.txt

It is a file for the 1st trait (litter size) with:  
 ITER: number of iteration  
 VE: chain for the variance of the error  
 RSD: chain for the Residual standard deviation (Square root of VE)  
 MEAN: chain for the general Mean of the model  
 ef1lev1: chain for the level 1 of the “fixed” effect 1  
 ef1lev2: chain for the level 2 of the “fixed” effect 1

## PROGRAM PostGibbs

### INPUT

#### Chain.txt

It is a file with two columns, the first one is the number of iteration and the second the MCMC chain to be analyzed. If the MCMC result comes from the BUGS program, we have several chains concatenated; for example, the first chain goes from iteration 1 to 5000, the second chain from iteration 5001 to 10000 and so on.

We will create a file taking ITER and a new column with the contrast [ef1lev1 – ef1lev2] from Results.txt. This contrast compares line 1 and line 2 for survival.

1	0.0193
2	0.015
3	-0.0077
4	-0.0043
5	0.0133
6	0.0068
7	0.0069
8	0.0032
9	0.0294
10	0.0033
11	0.0198
12	0.0311

**We will apply the PostGibbs program as before**

### 3. ANIMAL MODEL

EXAMPLE: A divergent selection experiment on litter size has been performed and we have data of litter size from the High and Low lines. We want to compare both lines at constant ovulation rate; i.e., we want the response to selection not dependent on possible differences in ovulation rate between lines. Data have been measured in two seasons.

#### MODEL

$$y_{ikm} = \mu + \text{Line}_i + \text{Season}_k + b \cdot \text{OvulationRate}_{ikm} + e_{ikm}$$

#### TRAITS

##### Litter size

#### “FIXED” EFFECTS

##### Line

Line 1 selected for high litter size  
Line 2 selected for low litter size

##### Season

Season 1 = Autumn  
Season 2 = Summer

#### COVARIATE

##### Ovulation rate

Number of data = 77  
Number of animals in the pedigree = 172

#### DATA FILE

Animal	Line	Season	OvulationRate	LitterSize
1	1	1	14	10
2	1	2	16	13
3	1	1	14	13

4	1	2	15	14
5	1	2	14	12
6	2	1	13	11
7	1	1	15	15
8	1	1	11	9
9	1	2	12	6
10	1	2	17	10
11	1	1	10	10
12	1	1	12	10

### PEDIGREE FILE

Animal	Sire	Dam
1	95	126
2	83	139
3	99	140
4	88	144
5	100	145
6	86	129
...	...	...
...	...	...
78	0	0
79	0	0
80	0	0
81	0	0
...	...	...

### PROCEDURE

#### **PROGRAM Animal**

#### INPUT

#### **Data.txt**

It is the data file, in txt format with tabs, with no head in the first line:

1	1	1	14	10
2	1	2	16	13
3	1	1	14	13
4	1	2	15	14
5	1	2	14	12
6	2	1	13	11
7	1	1	15	15
8	1	1	11	9
9	1	2	12	6
10	1	2	17	10
11	1	1	10	10
12	1	1	12	10

**Pedi.txt**

It is the pedigree file, in txt format with tabs, with no head in the first line:

1	95	126
2	83	139
3	99	140
4	88	144
5	100	145
6	86	129

**Files.txt**

It is a file needed by the program. Do not modify it.

## PROGRAM

**Animal**

Enter the total number of data

77

Enter the total number of animals in the pedigree file

172

Enter the number of 'fixed' effects

2

Enter the number of levels of 'fixed' effect 1

2

Enter the number of levels of 'fixed' effect 2

2

Enter the number of covariates (0.1.2.....)

1

Would you like to introduce the random seed (1=Yes. 2=No)

1

Enter a random seed

1234

Do you want to establish the MCMC characteristics? (1=Yes. 2=No)

1

Enter chain length

60000

Enter burn-in

10000

Enter lag between samples

10

## OUTPUT

**Results1.txt**

It is a file for the 1st trait (litter size) with:

ITER: number of iteration

VE: chain for the variance of the error

RSD: chain for the Residual standard deviation (Square root of VE)  
 VA: chain for the Additive variance  
 h2: chain for the heritability  
 MEAN: chain for the general Mean of the model  
 ef1lev1: chain for the level 1 of the "fixed" effect 1  
 ef1lev2: chain for the level 2 of the "fixed" effect 1  
 ef2lev1: chain for the level 1 of the "fixed" effect 2  
 ef2lev2: chain for the level 2 of the "fixed" effect 2  
 cov1: chain for the covariate

## PROGRAM PostGibbs

INPUT

### Chain.txt

It is a file, with no heading, with two columns, the first one is the number of iteration and the second the MCMC chain to be analyzed. If the MCMC result comes from the BUGS program, we have several chains concatenated; for example, the first chain goes from iteration 1 to 5000, the second chain from iteration 5001 to 10000 and so on.

We will create a file taking ITER and h2 from Results.txt

1	0.9381
2	0.8566
3	0.7877
4	0.7755
5	0.8497
6	0.8059
7	0.8074
8	0.6533
9	0.7271
10	0.8518
11	0.7634
12	0.5768
13	0.5488

### Figures for inference

We have to decide which the relevant value for the trait is. For heritability we decide that  $k=0.10$  using economical arguments. This permits to calculate the probability of relevance  $P>k$  or  $P<k$ . Notice that  $P>0$  will be always 1.

PROGRAM

### PostGibbs

Enter a chain from BUGS

Fist iteration

1

Last iteration  
5000  
Enter k for  $P < k$  or  $P > k$   
0.1  
Enter the probability for the HPD interval  
0.95  
Do you want to calculate the probability of an interval  $[a,b]$ ? (Y=1, N=0)  
0

## OUTPUT

### Features.txt

Mean = 0.6872

Median = 0.7391

( $P > 0.1000$ ) = 0.98

( $P < 0.1000$ ) = 0.02

HPD 0.95 = [ 0.2164 0.9977 ]

k for the interval  $[k, +\infty)$  with probability 95% = 0.2161

Notice that here the interval is  $[k, 1]$  instead of  $[k, +\infty)$ . We know that  $h^2$  is at least 0.2161 with a probability of 95%.

MCse = 0.00831

Geweke Z-Score = 0.3584

## 4. REPEATABILITY MODEL

EXAMPLE: A divergent selection experiment on litter size has been performed and we have data of neonatal survival (number of born alive / litter size) from several parities of does of the High and Low lines. We want to compare both lines. Former analyses suggest that we do not have a parity effect, but we can still have a “random” effect of doe which includes genetics and environmental permanent effects.

MODEL

$$y_{ikm} = \mu + \text{Line}_i + a_{ik} + p_{ik} + e_{ikm}$$

TRAITS

**Neonatal survival**

“FIXED” EFFECTS

**Line**

Line 1 selected for high litter size

Line 2 selected for low litter size

“RANDOM” EFFECTS

**Breeding value (genetic additive value)  $a_{ik}$**

**Permanent environmental effect  $p_{ik}$**

Number of data = 144

Number of animals in the pedigree = 180

Number of animals with data = 73

DATA FILE

Doe	Line	Survival
1	1	0.857143
1	1	0.2
2	1	1
2	1	1
3	1	1

3	1	1
4	1	0.5
4	1	0.625
5	2	0.833333
5	2	1
6	2	1

## PEDIGREE FILE

Animal	Sire	Dam
1	81	118
2	74	132
3	87	134
4	85	140
5	76	115
6	91	117
...	...	...
...	...	...
74	0	0
75	0	0
76	0	0
77	0	0
...	...	...

## PROCEDURE

### **PROGRAM Repeatability**

#### INPUT

#### **Data.txt**

It is the data file, in txt format with tabs, with no head in the first line:

1	1	0.85
1	1	0.20
2	1	1
2	1	1
3	1	1
3	1	1
4	1	0.50
4	1	0.62
5	2	0.83
5	2	1

**Pedi.txt**

It is the pedigree file, in txt format with tabs, with no head in the first line:

1	81	118
2	74	132
3	87	134
4	85	140
5	76	115
6	91	117

**Files.txt**

It is a file needed by the program. Do not modify it.

## PROGRAM

**Repeatability**

Enter the total number of data

144

Enter the total number of animals in the pedigree file

180

Enter the total number of animals with data

73

Enter the number of 'fixed' effects

1

Enter the number of levels of 'fixed' effect

2

Enter the number of covariates (0.1.2.....)

0

Would you like to introduce the random seed (1=Yes. 2=No)

1

Enter a random seed

4321

Do you want to establish the MCMC characteristics? (1=Yes. 2=No)

1

Enter chain length

60000

Enter burn-in

10000

Enter lag between samples

10

## OUTPUT

**Results.txt**

It is a file for the 1st trait (litter size) with:

ITER: number of iteration

VE: chain for the variance of the error

RSD: chain for the Residual standard deviation (Square root of VE)

VA: chain for the additive variance

VPe: chain for the permanent variance  
MEAN: chain for the general Mean of the model  
ef1lev1: chain for the level 1 of the “fixed” effect 1  
ef1lev2: chain for the level 2 of the “fixed” effect 1

## **PROGRAM PostGibbs**

INPUT

### **Chain.txt**

It is a file with two columns, the first one is the number of iteration and the second the MCMC chain to be analyzed. If the MCMC result comes from the BUGS program, we have several chains concatenated; for example, the first chain goes from iteration 1 to 5000, the second chain from iteration 5001 to 10000 and so on.

We will create a file taking ITER and a new column for the heritability with  $VA/(VA+VPe)$  from Results.txt

1	0.818808269
2	0.839855072
3	0.80530303
4	0.759425494
5	0.669673055
6	0.655364091
7	0.732734419
8	0.672839506

**We will apply the PostGibbs program as before**

## 5. RUNNING “BUGS”

EXAMPLE : We want to compare carcass traits of two breeds of rabbits, one of them selected for litter size and the other one of the for growth rate. We want to do this at constant liveweight.

### MODEL

$$y_{ikm} = \mu + \text{Line}_i + b \cdot \text{LiveWeight}_{ikm} + e_{ikm}$$

### “FIXED” EFFECTS

#### Line

Line A selected for litter size = 1

Line R selected for growth rate = 2

### COVARIATE

#### Liveweight

N. data: 320

### PRIORS

BUGS has a wide range of possibilities for defining priors. Here we put a bounded flat prior for the mean. The bounds should be defined for each trait according to our prior knowledge of the trait. An unorthodox (although common) procedure is to define the prior from the data, taken the average plus minus ten standard deviations. We make a loop in the program to assign to each level of the effect the same flat prior we assigned to the mean.

We assign a normal prior with zero mean and a high variance to the covariance. It is almost a flat prior, but it is a proper prior. BUGS works with *precision* (the inverse of the variance) instead of variance, thus we assign a precision of  $10^{-6}$  corresponding to a variance of  $10^6$ .

We consider the residuals to be normally distributed with zero mean and the same variance. We assign a bounded uniform prior to the variance. As before, we take the unorthodox decision of using the sample to determine the bound. We will say that the maximum value that a priori can take the variance is four times the variance of the sample.

### RESTRICTION

We make a conventional restriction, all the levels of the effect will sum zero.

# PROGRAM FILE

```

#   MODIFY ONLY RED TYPED

#   ;CAREFUL!: BUGS RECOGNIZES CAPITAL LETTERS

# _____

#   DATA DESCRIPTION

    data
    list(N=320,L=2)
    model
    {

#   N is the total number of data
#   L is the number of levels of the first effect
#   If there are more effects, other letters must be added, for example,
list(N=320,L=2,M=4)

# _____

#   PRIOR FOR THE MEAN

#   We assign a uniform prior with bounds (a1,a2) that are the sample mean  $\pm$  10 sd

    a2<- mean( trait [])+(sd( trait [])*10)
    a1<- - a2
    m~ dunif(a1 , a2)

# _____

#   PRIOR FOR THE EFFECT

#   We assign a uniform prior for each level with the same bounds as the mean

    for (i in 1: L -1) {
    effect [i] ~ dunif(a1 , a2)}

#   If there are more effects, this part should be repeated USING ANOTHER LETTER instead of L
#   There should be a correspondence between this letter and the letter used in the data description

# _____

# _____

#   PRIOR FOR THE COVARIATE

#   We assign a vague proper prior (a Normal with large variance: 1,000,000).

```

```

covariate1 ~ dnorm(0,1.0E-06)

# Repeat this part for each covariate
# _____

# MODEL

# We add a restriction: The sum of the level estimates should be zero

effect[_L] <- -sum( effect [1: L -1])

# Data are normally distributed with mean mu[i] and variance var for each level of
the effect
# Variance has a bounded flat prior. The bound is four times the sample variance

b <- sd(trait [])*sd(trait [])*4
var ~ dunif(0,b)
tau <- 1/var

for (i in 1:N) {
trait [i] ~ dnorm(mu[i] , tau)
mu[i] <- m+ effect1 [ef1[i]] + covariate1 *x1[i] }
}

# Add effect2, effect3, covariate2, covariate3, etc., by naming them changing "1" by
"2", etc.

# _____

```

## DATA FILE

```

ef1[] x1[] trait[]
1 2.200 1.9629
1 2.000 4.3621
2 1.900 3.4896
2 2.030 2.6172
2 2.430 3.2715

```

```

# Add ef2[] x2[] etc. when having more effects or covariates, for example ef1[]
ef2[] x1[] x2[]

```

```

END

```

# PROCEDURE

- Open PROGRAM FILE in BUGS clicking on the Menu

**File**

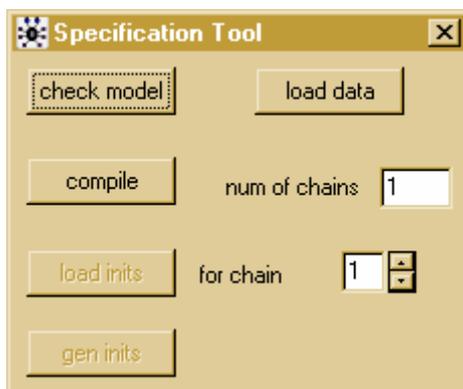
**Open**

- Select **model** in PROGRAM FILE

- Click on the Menu

**Model**

**Specification**



- Click on **check model**
- Select **list** in the PROGRAM FILE
- Click on **load data**
- Open DATA FILE clicking on the Menu **File** **Open**
- Select the first row containing **ef1[]** **x1[]**
- Click on **load data**  
**compile**  
**gen inits**

- Click on the Menu

**Inference**

**Samples**

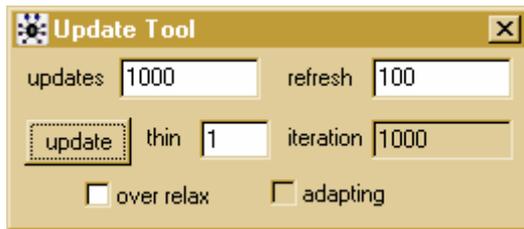


- Write name of **effect1** in **node**
- Click **set**
- Repeat with the other effects and covariates
- Write 1001 in **beg**  
Write 50000 in **end**
- Click on **trace**

- Click on the Menu

**Model**

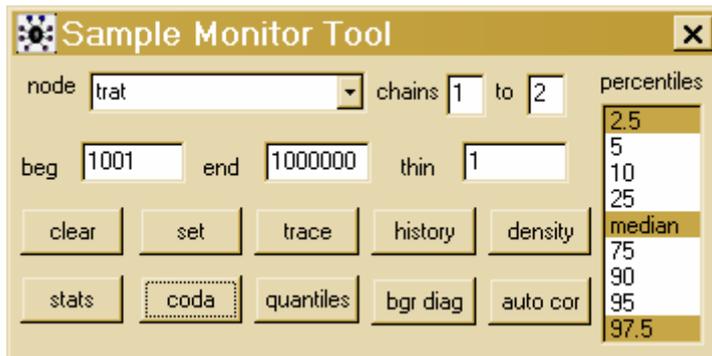
**Update**



- Write 10000 in **updates**

- Click on the Menu

**Inference**  
**Samples**



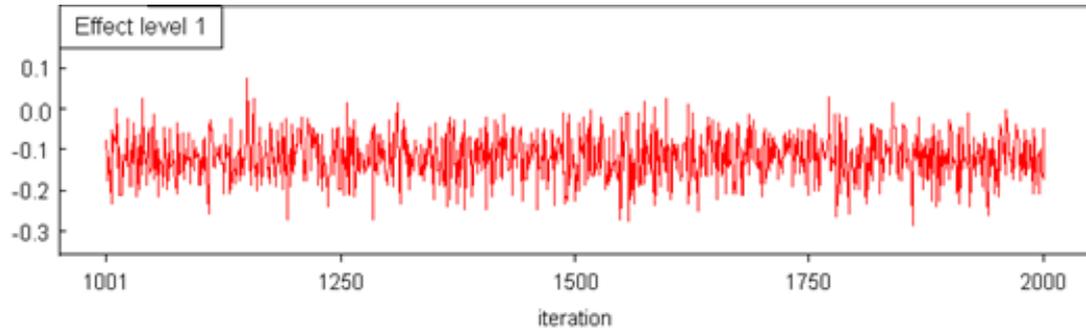
- Click on

**history**  
**density**  
**stats**  
**coda**

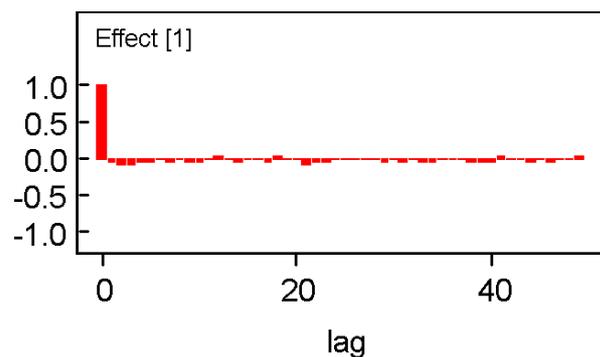
- Go to the file **CODA for chain 1** and save it as "Plain text" clicking in the Menu  
**File**  
**Save as**
- Go to the file **CODA index** and annotate the beginning of each chain
- Make inferences with **PostGibbs** function

## OUTPUT

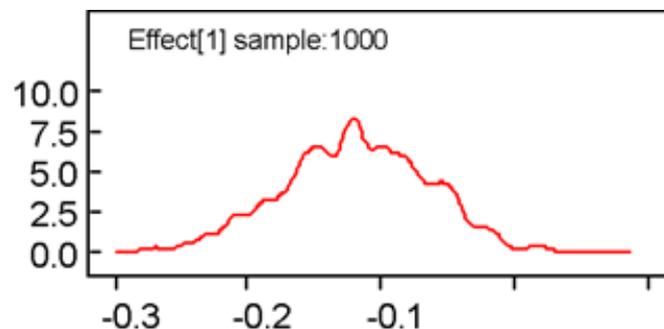
**Graph of the chain:** We get one graph for each level of the effect, for the mean and the covariate and for the residual variance. We see that in the chain convergence is achieved.



**Autocorrelation:** Shows the correlation between two consecutive samples, and between one sample and the next third, fourth, fifth, samples, etc. (Figure 4.9)



**Marginal posterior distribution:** Shows a graph of the marginal posterior distribution estimated from a MCMC chain. We get a similar graph for each unknown.



**Statistics:** This file gives the features of the marginal posterior distribution of each level of the effect and the number of samples of the chain.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
sexo[1]	-0.1194	0.05492	0.001217	-0.2296	-0.1189	-0.01595	1001	1000
sexo[2]	0.1194	0.05492	0.001217	0.01723	0.119	0.2299	1001	1000

**Chains:** BUGS gives the chains consecutively. Samples are numbered. First column gives the number of the sample and second column the value of the sample. First, the program says in the CODA FILE where the samples of each marginal posterior distribution start and finish. Here we find that the samples for the first level of the effect start in row 1 and finish in row 1000, then the samples for the second level of the effect start in row 1001 and finish in row 2000.

#### CODA FILE

```
Effect[1] 1 1000
Effect[2] 1001 2000
```

#### Chains

```
1001 -0.08111
1002 -0.1228
1003 -0.1794
1004 -0.1428
1005 -0.1172
1006 -0.2291
1007 -0.05237
1008 -0.1161
1009 -0.09516
1010 -0.06668
1011 -0.08592
1012 -0.001962
1013 -0.06813
1014 -0.1442
.....
.....
```

## FROM BUGS MANUAL

### UPDATE TOOL

**updates:** number of MCMC updates to be carried out.

**refresh:** the number of updates between redrawing the screen.

**thin:** the samples from every  $k^{\text{th}}$  iteration will be stored, where  $k$  is the value of **thin**. Setting  $k > 1$  can help to reduce the autocorrelation in the sample, but there is no real advantage in thinning except to reduce storage requirements and the cost of handling the simulations when very long runs are being carried out.

**update:** Click to start updating the model. Clicking on *update* during sampling will pause the simulation after the current block of iterations, as defined by *refresh*, has been completed; the number of updates required can then be changed if needed. Clicking on update again will restart the simulation. This button becomes active when the model has been successfully compiled and given initial values.

**iteration:** shows the total number of iterations stored after thinning – not the actual number of iterations carried out. In this respect *updates* represents the required number of samples rather than MCMC updates: for example, if 100 samples are requested (via *updates* = 100) and *thin* is set equal to 10, then  $10 * 100 = 1000$  iterations will actually be carried out, of which 100 (every 10th) will be stored.

### SAMPLE MONITOR TOOL

**node:** The variable of interest must be typed in this text field. If the variable of interest is an array, slices of the array can be selected using the notation `variable[lower0:upper0, lower1:upper1, ...]`. The buttons at the bottom of the dialog act on this variable. A star '\*' can be entered in the node text field as shorthand for all the stored samples.

**beg** and **end:** numerical fields are used to select a subset of the stored sample for analysis.

**thin:** numerical field used to select every  $k^{\text{th}}$  iteration of each chain to contribute to the statistics being calculated, where  $k$  is the value of the field. Note the difference between this and the thinning facility on the [Update Tool](#) dialog box: when thinning via the *Update Tool* we are *permanently* discarding samples as the MCMC simulation runs, whereas here we have already generated (and stored) a suitable number of (posterior) samples and may wish to discard some of them only temporarily. Thus, setting  $k > 1$  here will not have any impact on the storage (memory) requirements of processing long runs; if you wish to reduce the number of samples actually stored (to free-up memory) you should thin via the *Update Tool*.

**chains . to .:** can be used to select the chains which contribute to the statistics being calculated.

**clear:** removes the stored values of the variable from computer memory.

**set:** must be used to start recording a chain of values for the variable.

**trace:** plots the variable value against iteration number. This trace is dynamic, being redrawn each time the screen is redrawn.\*

**history:** plots out a complete trace for the variable.\*

The next six buttons will be greyed out if the MCMC simulation is in an adaptive phase.

**density:** plots a smoothed kernel density estimate for the variable if it is continuous or a histogram if it is discrete.\*

**auto cor:** plots the autocorrelation function of the variable out to lag-50. The values underlying these can be listed to a window by double-clicking on the plot followed by *ctrl-left-mouse-click*: if multiple chains are being simulated the values for each chain will be given.\*

**stats:** produces summary statistics for the variable, pooling over the chains selected. The required percentiles can be selected using the *percentile* selection box. The quantity reported in the MC error column gives an estimate of  $\sigma / N^{1/2}$ , the Monte Carlo standard error of the mean. The batch means method outlined by Roberts (1996; p.50) is used to estimate  $\sigma$ .

**coda:** dumps out an ascii representation of the monitored values suitable for use in the CODA S-Plus diagnostic package. An output file for each chain is produced, corresponding to the `.out` files of CODA, showing the iteration number and value (to four significant figures). There is also a file containing a description of which lines of the `.out` file correspond to which variable – this corresponds to the CODA `.ind` file. These can be named accordingly and saved as text files for further use. (Care may be required to stop the Windows system adding a `.txt` extension when saving: enclosing the required file name in quotes should prevent this.)

**quantiles:** plots out the running mean with running 95% confidence intervals against iteration number.\*

**bgr diag:** calculates the Gelman-Rubin convergence statistic, as modified by Brooks and Gelman (1998). The width of the central 80% interval of the pooled runs is green, the average width of the 80% intervals within the individual runs is blue, and their ratio  $R$  (= pooled / within) is red - for plotting purposes the pooled and within interval widths are normalised to have an overall maximum of one. The statistics are calculated in bins of length 50:  $R$  would generally be expected to be greater than 1 if the starting values are suitably over-dispersed. Brooks and Gelman (1998) emphasise that one should be concerned both with convergence of  $R$  to 1, and with convergence of both the pooled and within interval widths to stability.

The values underlying these plots can be listed to a window by double-clicking on the plot followed by *ctrl-left-mouse-click* on the window.\*

\* See [WinBUGS Graphics](#) for details of how to customize these plots.