



# GenBUGS & R - Running external programs from GenStat

Darren Murray

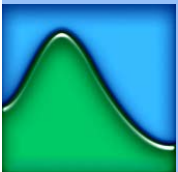
VSN International, 5 The Waterhouse,  
Waterhouse Street, Hemel Hempstead, UK

Wageningen, 18<sup>th</sup> June 2008



# Bayesian Methods

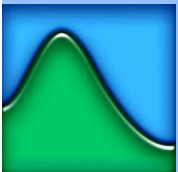
- Another philosophy
  - Currently applied in a wide range of areas
  - Popular alternative to traditional methods
- Bayesian computation
  - Simulation methods applied to Bayesian framework
- Markov Chain Monte Carlo (MCMC)
  - Iterative method for drawing random samples from posterior distribution
  - Fit wide range of (complex) models including modelling latent variables & estimation of hierarchical models
- Many freeware applications now available for Bayesian analysis



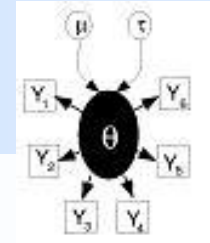
# The BUGS project



- **Bayesian inference Using Gibbs Sampling**
  - Software for Bayesian analysis of statistical models using Markov Chain Monte Carlo (MCMC) methods
  - Run by MRC Biostatistics Unit, Cambridge
  - [www.mrc-bsu.cam.ac.uk/bugs/](http://www.mrc-bsu.cam.ac.uk/bugs/)
  - Several different versions
- **'Classic' BUGS**
  - Developed by MRC Biostatistics Unit in 1989
  - Command based
  - Available on several different platforms
  - No longer developed



# BUGS

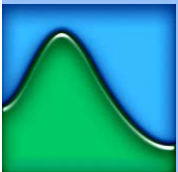


- WinBUGS

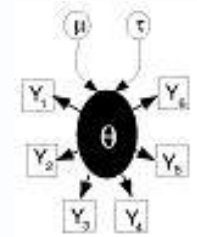
- Developed by MRC Biostatistics Unit and Imperial College School of Medicine, St Mary's, London
- Most widely known version
- Windows application uses menus and commands
- Add-on packages available (GeoBUGS, PKBUGS)

- OpenBUGS

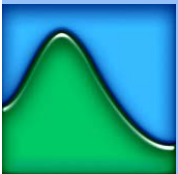
- Open-source for the core BUGS code with a variety of interfaces
- Developed at University of Helsinki
  - <http://www.rni.helsinki.fi/openbugs>
- Provides a WinBUGS interface
  - Currently add-on packages do not work
- Runs under linux as LinBUGS
- Now the major developmental platform



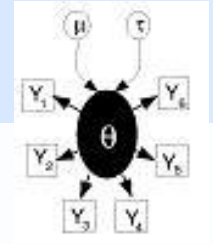
# Health Warning!



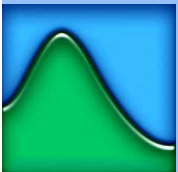
- BUGS assumes the user is familiar with Bayesian analysis
- It also assumes the user recognises the potential importance of prior distributions
- MCMC is less robust than analytical statistical methods
- No protection for fitting inappropriate models
- Poor initial values will result in failure!



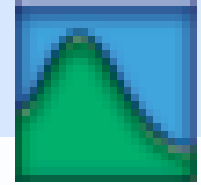
# Steps to run BUGS



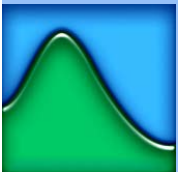
1. Specify the model
2. If data exists, then load data
3. Compile model (and data)
4. Specify initial values
5. Generate any other required initial values
6. Update to run burn-in.
7. Specify nodes to be monitored
8. Run MCMC
9. View results



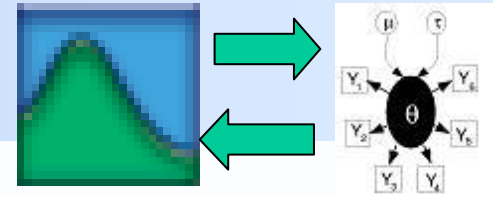
# Why use GenStat?



- Exporting data
  - GenStat has much better data management facilities
- Easy to change data values
  - More convenient than WinBUGS where you need to use copy & paste
- Can process a series of data sets
- Import data to GenStat for further analysis
  - Graphical displays and posterior simulations
  - Can write programs in GenStat to extend facilities in GenBUGS



# GenBUGS

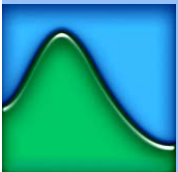


- GenBUGS

- set of procedures for running BUGS from GenStat in batch mode
- BGXGENSTAT, BGIMPORT & BG PLOT
- Provides tools for importing results, graphical display and convergence statistics

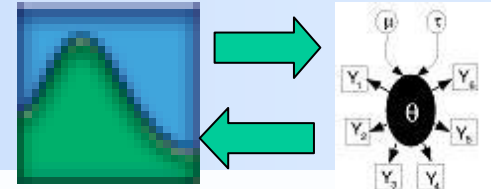
- Compatibility

- WinBUGS version 1.4
- OpenBUGS version 2.2 in Windows
- Plan to extend to linux with LinBUGS

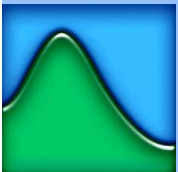




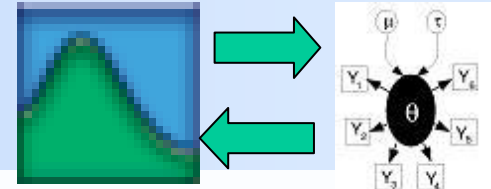
# Running GenBUGS



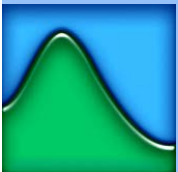
- Run using the BGXGENSTAT procedure
  - A wrapper for different steps to run BUGS
- Steps
  - Writes the data and initial value files used by BUGS into the working directory
  - Writes a script file to the BUGS executable folder using options from the procedure to load, compile and run
  - Runs BUGS using the executable file
  - Imports results from files saved by BUGS



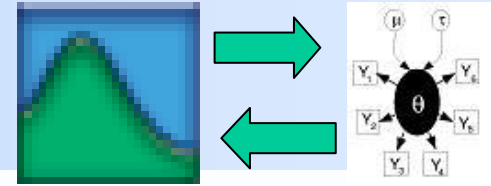
# BGXGENSTAT



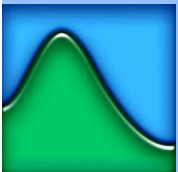
- The model is supplied in a file using BUGS script
  - MODELFILE option
- Data are supplied in a pointer and the names for each variable in a text
  - DATA and IDATANAMES options
- Names of nodes to be monitored supplied using MONITOR option
- Initial values are supplied in a list of pointers for each chain
  - The initial value names are specified using INAMES option
  - If left blank BUGS will generate values automatically
  - In practice BUGS can fail when reasonable initial values not specified



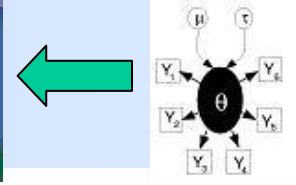
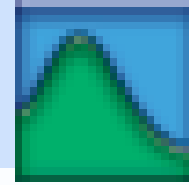
# BGXGENSTAT



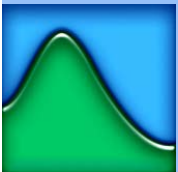
- The location and name of the BUGS executable must be supplied
  - WPATH & WEXE options
- The length of the burn-in and number of iterations (chains) are supplied with NBURNIN and NCHAINS options
- The number of simulation draws is specified by the NSAMPLES option
- The Deviance Information Criterion (DIC) can be computed using the DIC option
  - the posterior mean of the deviance plus  $P_D$ , the estimated effective number of parameters in the posterior distribution



# Importing results



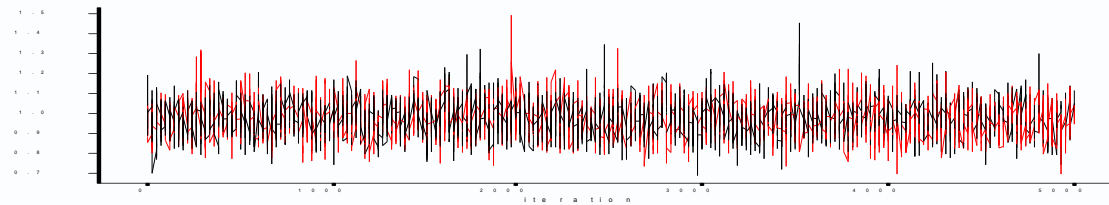
- WinBUGS can export results for monitored nodes
  - CODA option of BGXGENSTAT can be used to save data
  - Consist of an index file and a file for each chain
  - Results can be used for graphical displays and convergence diagnostics
- BGIMPORT procedure
  - Imports the results for monitored values into GenStat
  - Input requires index filename and prefix for output files
  - Stores results in pointers containing variates of the monitored values and text with the node names



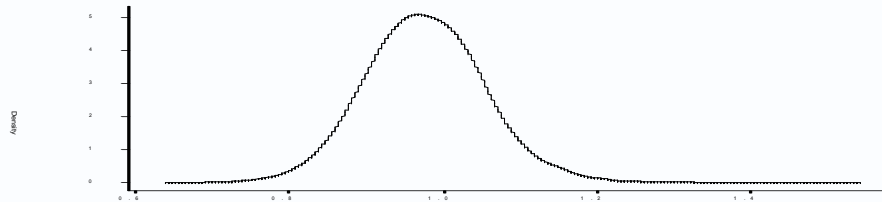
# Graphical displays

- BGPLOT procedure

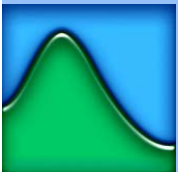
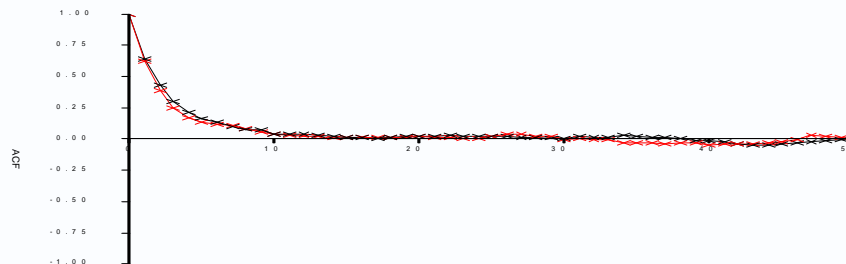
- Data are supplied as pointers as stored by BGIMPORT
- Trace plot of each monitored node



- Density plot using kernel density smooth

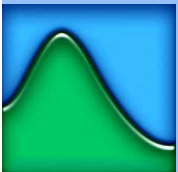
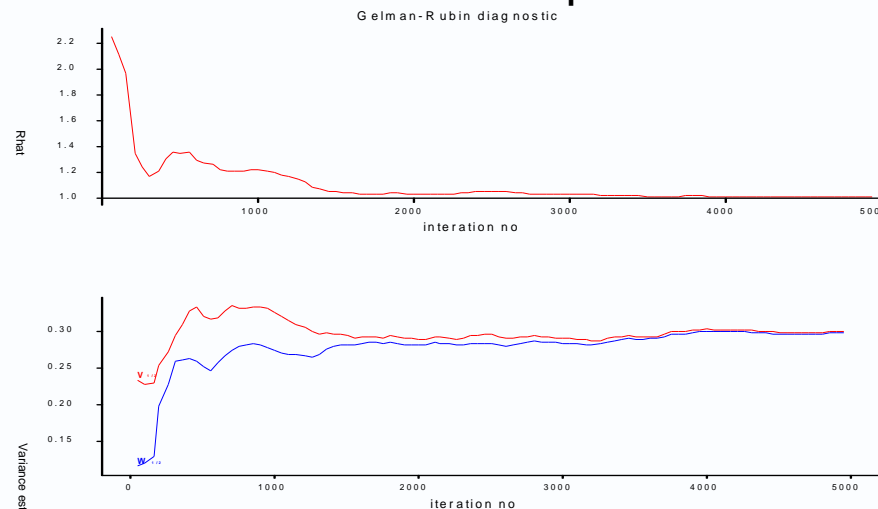


- Autocorrelation plot



# Convergence diagnostic

- Gelman-Rubin 'Potential scale reduction factor' diagnostic
  - Compares between and within variances of multiple chains
  - Approximate convergence when close to 1
- BGPlot procedure
  - Plots the psrf as number of iterations increases (Brooks & Gelman - 1998)
  - Plots the between and within-sequence variance



# Example – salamander data

Spreadsheet [Salamander.gsh]\*

Row	Season	Experiment	TypeM	TypeF	Cross	Male	Female	Mate
1	Summer	1	R	R	RR	1	1	1
2	Summer	1	W	R	RW	14	1	1
3	Summer	1	R	R	RR	5	1	1
4	Summer	1	W	R	RW	11	1	0
5	Summer	1	R	R	RR	4	1	1
6	Summer	1	W	R	RW	15	1	1
7	Summer	1	R	R	RR	5	2	1
8	Summer	1	W	R	RW	15	2	1
9	Summer	1	R	R	RR	3	2	1
10	Summer	1	W	R	RW	13	2	1
11	Summer	1	R	R	RR	1	2	1
12	Summer	1	W	R	RW	12	2	1
13	Summer	1	R	R	RR	2	3	1
14	Summer	1	W	R	RW	11	3	0
15	Summer	1	R	R	RR	1	3	1
16	Summer	1	W	R	RW	14	3	1
17	Summer	1	R	R	RR	3	3	1
18	Summer	1	W	R	RW	13	3	1
19	Summer	1	R	R	RR	4	4	1
20	Summer	1	W	R	RW	12	4	1
21	Summer	1	R	R	RR	2	4	1
22	Summer	1	W	R	RW	15	4	0
23	Summer	1	R	R	RR	5	4	1
24	Summer	1	W	R	RW	14	4	1

- LNP §6.6.1, McC & N §14.5
- 3 experiments each mating 20 male & 20 female salamanders of 2 types (whiteside and rough butt), paired 6 times with their own and the other type (so 120 observations per experiment)
- standard dataset to demonstrate deficiencies of GLMM algorithms

..

# Example - GLMM

- Salamander mating data set
  - McCullagh & Nelder (1989), Lee, Nelder & Pawitan (1996)

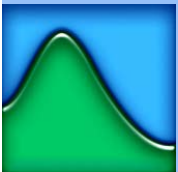
$$p_{ijk} \sim \text{binomial}(\text{Mate}_{ijk}, n_{ijk})$$

$$\begin{aligned} \text{logit}(\text{Mate}_{ijk}) = & a_0 + a_1 \text{Trtf}_{ijk} + a_2 \text{Trtm}_{ijk} + a_{12} (\text{Trtf} \cdot \text{Trtm})_{ijk} \\ & + b^f_{ik} + b^m_{jk} \end{aligned}$$

for  $i, j = 1 \dots 20$  and  $k = 1, 2, 3$

$$b^f_{ik} \sim \text{Normal}(0, \sigma^2_f)$$

$$b^m_{jk} \sim \text{Normal}(0, \sigma^2_m)$$

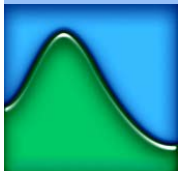




# Example – salamander data

Estimates from the mean model				Gibbs	MCEM	
	estimate	s.e.	t(356)	1.03	1.02	
constant	1.043	0.3932	2.653	-3.01	-2.96	
TypeF W	-3.005	0.5223	-5.754	-0.69	-0.70	
TypeM W	-0.729	0.4529	-1.609	3.74	3.63	
TypeF W .TypeM W	3.713	0.5569	6.667			
Estimates from the dispersion model						
Estimates of parameters						
Parameter	estimate	s.e.	t(*)	antilog of estimate	Gibbs	MCEM
lambda Female	0.318	0.272	1.17	1.374	1.22	1.18
lambda Male	0.187	0.280	0.67	1.206	1.17	1.11
..						

- improved Laplace for dispersion and mean models:  
MLAPLACE=1; DLAPLACE=1



# Running R from GenStat

- Works similarly to running WinBUGS
  - Runs a R in batch mode by creating a basic R script
- Allows data to be passed between applications using R data frames
  - Reference in R in usual way
- Procedure RXGENSTAT
  - Set path to R executable
  - Set working directory
  - Supply a R script as text or in file
  - Allows one data frame for import/export

