Performing Bayesian analysis in Stata using WinBUGS

Tom Palmer, John Thompson & Santiago Moreno

Department of Health Sciences,
University of Leicester, UK

13th UK Stata Users Group Meeting,
10 September 2007
Outline

1. The Bayesian approach & WinBUGS
2. The winbugsfromstata package
3. How to run an analysis
4. Summary & developments
The Bayesian approach

Bayes Theorem

\[ \text{Posterior} \propto \text{Likelihood} \times \text{prior} \]
Bayes Theorem

\[ \text{Posterior} \propto \text{Likelihood} \times \text{prior} \]

- Direct probability statements - not frequentist - subjective
Bayes Theorem

\[ \text{Posterior} \propto \text{Likelihood} \times \text{prior} \]

- Direct probability statements - not frequentist - subjective
- Complex posterior marginal distributions - estimation via simulation
The Bayesian approach

Bayes Theorem

\[ \text{Posterior} \propto \text{Likelihood} \times \text{prior} \]

- Direct probability statements - not frequentist - subjective
- Complex posterior marginal distributions - estimation via simulation
- Markov chain Monte Carlo (MCMC) methods
- Bayesian statistics using Gibbs sampling
WinBUGS

- Bayesian statistics using Gibbs sampling

- MRC Biostatistics unit
  http://www.mrc-bsu.cam.ac.uk/bugs
Bayesian statistics using Gibbs sampling

MRC Biostatistics unit
http://www.mrc-bsu.cam.ac.uk/bugs

Health Economics, Medical Statistics
Bayesian statistics using Gibbs sampling

MRC Biostatistics unit
http://www.mrc-bsu.cam.ac.uk/bugs

Health Economics, Medical Statistics

Disadvantages: data management, post-processing of results, graphics
The `winbugsfromstata` package

- Stata interface to WinBUGS [Thompson et al., 2006]
  http://www2.le.ac.uk/departments/health-sciences/extranet/BGE/genetic-epidemiology/gedownload/information
The winbugsfromstata package

Manual
A guide to the programs presented in the style of the Stata Reference Manuals is available for download.

Examples
A number of worked examples demonstrating the use of the programs are provided below.

Oxford example

Blocker example

Ado-files and help files for manual installation
The ado and help files are available for download. Save the files on the adopath.

wbfiles
How to run an analysis

1. Read in data from a BUGS example
   - `wbdecode`

2. Data preparation
   - `wbarray`, `wbdata`, `wbscalar`, `wbstructure`, `wvvector`
   - TEXT FILE

3. Define WinBUGS script
   - `wbscript`
   - TEXT FILE

4. Initial values
   - data commands
   - TEXT FILE

5. Run model
   - `wbrun`

6. WinBUGS

7. Read in WinBUGS results
   - `wbcoa`

8. Summarizing MCMC results
   - `wbstats`, `wbdensity`, `wbdic`, `wbhull`

9. Convergence & mixing
   - `wbac`, `wbbgr`, `wbgeweke`, `wbintervals`, `wbsection`, `wbtrace`
help winbugs

Title

winbugs — help on available routines for running WinBUGS

Description

This help file describes the commands available for running WinBUGS from within Stata. There is no executable command winbugs.

The files are:

- wbarray: writes data from Stata as a WinBUGS array
- wbdata: writes mixed data (scalars, vectors & structures) from Stata as a WinBUGS list
- wbdecode: writes scalars from Stata as a WinBUGS list
- wbstructure: writes data from Stata as a WinBUGS structure
- wbvector: writes data from Stata as a WinBUGS vector
- wbcoda: reads data from a WinBUGS coda file into Stata
- wbcodex: reads data from a WinBUGS list into Stata
- wbchn: runs a pre-prepared WinBUGS script file from within Stata
- wbcscript: writes & runs a WinBUGS script file from within Stata
- wbac: autocorrelation plots
- wbbbr: Brooks-Gelman-Rubin plot
- wbbgeweke: test of means for two sections of a chain
- wbbintervals: interval plots for sections of a chain
- wbbsection: density plots of subsections of a chain
- wbbtrace: trace (history) plot(s) of an MCMC run
- wbdensity: smoothed posterior density estimates
- wbbdic: read Deviance Information Criterion (DIC) statistics in a WinBUGS log-file into Stata
- wbbhull: contours for pairs of parameters
- wbbsumstat: summary statistics from an MCMC chain.
**Example analysis: Schools**

- Schools example [Goldstein et al., 1993], [Spiegelhalter et al., 2004]
Example analysis: Schools

- Schools example [Goldstein et al., 1993], [Spiegelhalter et al., 2004]
- Between-school variation in exam results from inner London schools
Example analysis: Schools

- Schools example [Goldstein et al., 1993], [Spiegelhalter et al., 2004]
- Between-school variation in exam results from inner London schools
- Standardized mean scores ($Y$) 1,978 pupils, 38 schools
Example analysis: Schools

- Schools example [Goldstein et al., 1993], [Spiegelhalter et al., 2004]
- Between-school variation in exam results from inner London schools
- Standardized mean scores ($Y$) 1,978 pupils, 38 schools
- LRT: London Reading Test, VR: verbal reasoning, Gender intake of school, denomination of school
Data for the Schools example
The model

- Hierarchical model; specified the mean and variance
The model

- Hierarchical model; specified the mean and variance

Model:

\[ Y_{ij} \sim N(\mu_{ij}, \tau_{ij}) \]

\[ \mu_{ij} = \gamma_1 j + \gamma_2 j LRT_{ij} + \gamma_3 j VR_{1ij} + \beta_1 LRT_{ij}^2 + \beta_2 VR_{2ij} \]

\[ + \beta_3 Girl_{ij} + \beta_4 Gsch_j + \beta_5 Bsch_j + \beta_6 CEsch_j + \beta_7 RCsch_j + \beta_8 Osch_j \]

\[ \log \tau_{ij} = \theta + \phi LRT_{ij} \]
model{
for(p in 1 : N){
  Y[p] ~ dnorm(mu[p], tau[p])
    + beta[8] * School.denom[p, 3]
  log(tau[p]) <- theta + phi * LRT[p]
  sigma2[p] <- 1 / tau[p]
  LRT2[p] <- LRT[p] * LRT[p]
}
}
min.var <- exp(-(theta + phi * (-34.6193))) # lowest LRT score = -34.6193
max.var <- exp(-(theta + phi * (37.3807))) # highest LRT score = 37.3807

# Priors for fixed effects:
  for (k in 1 : 8){
    beta[k] ~ dnorm(0.0, 0.0001)
  }
  theta ~ dnorm(0.0, 0.0001)
  phi ~ dnorm(0.0, 0.0001)

# Priors for random coefficients:
  for (j in 1 : M) {
    alpha[j, 1 : 3] ~ dmnorm(gamma[1:3 ], T[1:3 ,1:3 ])
    alpha1[j] <- alpha[j,1]
  }

# Hyper-priors:
  gamma[1 : 3] ~ dmnorm(mn[1:3 ], prec[1:3 ,1:3 ])
  T[1 : 3, 1 : 3 ] ~ dwish(R[1:3 ,1:3 ], 3)
Do-file for the example

// winbugsfromstata demo, 16august2007
cd "Z:/conferences/stata.users.uk.2007/schools"
wbdecode, file(Schoolsdata.txt) clear

wbscript, sav('c(pwd)'/script.txt, replace) ///
model('c(pwd)'/Schoolsmodel.txt) ///
data('c(pwd)'/Schoolsdata.txt) ///
inits('c(pwd)'/Schoolsinits.txt) ///
coda('c(pwd)'/out) ///
burn(500) update(1000) ///
set(beta gamma phi theta) dic ///
log('c(pwd)'/winbugslog.txt) ///
quit

wbrun , sc('c(pwd)'/script.txt) ///
win(Z:/winbugs/WinBUGS14/WinBUGS14.exe)

clear
set memory 500m
wbcoda, root(out) clear

wbstats gamma* beta* phi theta

wbtrace beta_1 gamma_1 phi theta
wbdensity beta_1 gamma_1 phi theta
wbac beta_1 gamma_1 phi theta
wbhull beta_1 beta_2 gamma_2, peels(1 5 10 25)

wbgeweke beta_1 gamma_1 phi theta

wbdic using winbugslog.txt
Tom Palmer (Leicester)  Running WinBUGS from Stata
Running WinBUGS from Stata
### wbstats output

```
. wbstats gamma* beta* phi theta

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>sem</th>
<th>median</th>
<th>95% CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>500</td>
<td>-0.715</td>
<td>0.103</td>
<td>0.0179</td>
<td>-0.715</td>
<td>(-0.951, -0.523 )</td>
</tr>
<tr>
<td>gamma_2</td>
<td>500</td>
<td>0.031</td>
<td>0.010</td>
<td>0.0005</td>
<td>0.031</td>
<td>(0.010, 0.052 )</td>
</tr>
<tr>
<td>gamma_3</td>
<td>500</td>
<td>0.967</td>
<td>0.105</td>
<td>0.0225</td>
<td>0.972</td>
<td>(0.750, 1.168 )</td>
</tr>
<tr>
<td>beta_1</td>
<td>500</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>(0.000, 0.000 )</td>
</tr>
<tr>
<td>beta_2</td>
<td>500</td>
<td>0.433</td>
<td>0.072</td>
<td>0.0099</td>
<td>0.435</td>
<td>(0.284, 0.576 )</td>
</tr>
<tr>
<td>beta_3</td>
<td>500</td>
<td>0.173</td>
<td>0.048</td>
<td>0.0031</td>
<td>0.172</td>
<td>(0.085, 0.271 )</td>
</tr>
<tr>
<td>beta_4</td>
<td>500</td>
<td>0.151</td>
<td>0.141</td>
<td>0.0230</td>
<td>0.164</td>
<td>(-0.156, 0.392 )</td>
</tr>
<tr>
<td>beta_5</td>
<td>500</td>
<td>0.091</td>
<td>0.105</td>
<td>0.0150</td>
<td>0.087</td>
<td>(-0.094, 0.318 )</td>
</tr>
<tr>
<td>beta_6</td>
<td>500</td>
<td>-0.279</td>
<td>0.183</td>
<td>0.0279</td>
<td>-0.290</td>
<td>(-0.618, 0.108 )</td>
</tr>
<tr>
<td>beta_7</td>
<td>500</td>
<td>0.170</td>
<td>0.105</td>
<td>0.0158</td>
<td>0.169</td>
<td>(-0.029, 0.380 )</td>
</tr>
<tr>
<td>beta_8</td>
<td>500</td>
<td>-0.109</td>
<td>0.209</td>
<td>0.0376</td>
<td>-0.124</td>
<td>(-0.485, 0.357 )</td>
</tr>
<tr>
<td>phi</td>
<td>500</td>
<td>-0.003</td>
<td>0.003</td>
<td>0.0002</td>
<td>-0.003</td>
<td>(-0.009, 0.003 )</td>
</tr>
<tr>
<td>theta</td>
<td>500</td>
<td>0.579</td>
<td>0.032</td>
<td>0.0016</td>
<td>0.579</td>
<td>(0.513, 0.649 )</td>
</tr>
</tbody>
</table>
```
. wbstats gamma* beta* phi theta

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>sem</th>
<th>median</th>
<th>95% CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma_1</td>
<td>500</td>
<td>-0.715</td>
<td>0.103</td>
<td>0.0179</td>
<td>-0.715</td>
<td>( -0.951, -0.523 )</td>
</tr>
<tr>
<td>gamma_2</td>
<td>500</td>
<td>0.031</td>
<td>0.010</td>
<td>0.0005</td>
<td>0.031</td>
<td>( 0.010, 0.052 )</td>
</tr>
<tr>
<td>gamma_3</td>
<td>500</td>
<td>0.967</td>
<td>0.105</td>
<td>0.0225</td>
<td>0.972</td>
<td>( 0.750, 1.168 )</td>
</tr>
<tr>
<td>beta_1</td>
<td>500</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0000</td>
<td>0.000</td>
<td>( 0.000, 0.000 )</td>
</tr>
<tr>
<td>beta_2</td>
<td>500</td>
<td>0.433</td>
<td>0.072</td>
<td>0.0099</td>
<td>0.435</td>
<td>( 0.284, 0.576 )</td>
</tr>
<tr>
<td>beta_3</td>
<td>500</td>
<td>0.173</td>
<td>0.048</td>
<td>0.0031</td>
<td>0.172</td>
<td>( 0.085, 0.271 )</td>
</tr>
<tr>
<td>beta_4</td>
<td>500</td>
<td>0.151</td>
<td>0.141</td>
<td>0.0230</td>
<td>0.164</td>
<td>( -0.156, 0.392 )</td>
</tr>
<tr>
<td>beta_5</td>
<td>500</td>
<td>0.091</td>
<td>0.105</td>
<td>0.0150</td>
<td>0.087</td>
<td>( -0.094, 0.318 )</td>
</tr>
<tr>
<td>beta_6</td>
<td>500</td>
<td>-0.279</td>
<td>0.183</td>
<td>0.0279</td>
<td>-0.290</td>
<td>( -0.618, 0.108 )</td>
</tr>
<tr>
<td>beta_7</td>
<td>500</td>
<td>0.170</td>
<td>0.105</td>
<td>0.0158</td>
<td>0.169</td>
<td>( -0.029, 0.380 )</td>
</tr>
<tr>
<td>beta_8</td>
<td>500</td>
<td>-0.109</td>
<td>0.209</td>
<td>0.0376</td>
<td>-0.124</td>
<td>( -0.485, 0.357 )</td>
</tr>
<tr>
<td>phi</td>
<td>500</td>
<td>-0.003</td>
<td>0.003</td>
<td>0.0002</td>
<td>-0.003</td>
<td>( -0.009, 0.003 )</td>
</tr>
<tr>
<td>theta</td>
<td>500</td>
<td>0.579</td>
<td>0.032</td>
<td>0.0016</td>
<td>0.579</td>
<td>( 0.513, 0.649 )</td>
</tr>
</tbody>
</table>

- **regress \( \gamma_2 \): 0.030, 95% C.I. (0.026, 0.034)**
. wbgeweke beta_1
Parameter: beta_1 first 10.0% (n=50) vs last 50.0% (n=250)
Means (se)  0.0003 (  0.0000)  0.0003 (  0.0000)
Autocorrelations  0.3736  0.4114
Mean Difference (se)  0.0000 (  0.0000) z =  1.030 p =  0.3031
Stata output

wbgeweke output

```
. wbgeweke beta_1
Parameter: beta_1 first 10.0% (n=50) vs last 50.0% (n=250)
Means (se)  0.0003 ( 0.0000)  0.0003 ( 0.0000)
Autocorrelations  0.3736  0.4114
Mean Difference (se)  0.0000 ( 0.0000)  z = 1.030  p = 0.3031
```

wb dic output

```
. wbdic using winbugslog.txt
DIC statistics 1

DIC  
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes  

<table>
<thead>
<tr>
<th></th>
<th>Dbar</th>
<th>Dhat</th>
<th>pD</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>4466.330</td>
<td>4393.470</td>
<td>72.861</td>
<td>4539.190</td>
</tr>
<tr>
<td>total</td>
<td>4466.330</td>
<td>4393.470</td>
<td>72.861</td>
<td>4539.190</td>
</tr>
</tbody>
</table>
```

Tom Palmer (Leicester)

Running WinBUGS from Stata
wbtrace output
Tom Palmer (Leicester)  Running WinBUGS from Stata
wbac output

Autocorrelations of beta_1

Autocorrelations of gamma_1

Autocorrelations of phi

Autocorrelations of theta

Bartlett's formula for MA(q) 95% confidence bands
wbhull output
WinBUGS - easy & flexible
Summary

- WinBUGS - easy & flexible
- `winbugsfromstata` - data preparation, analysis of MCMC output, graphics
Summary

- WinBUGS - easy & flexible
- `winbugsfromstata` - data preparation, analysis of MCMC output, graphics
- Prior distributions - *controversial*
WinBUGS - easy & flexible

*winbugsfromstata* - data preparation, analysis of MCMC output, graphics

Prior distributions - *controversial*

Check complex Stata models - *vague* prior distributions
WinBUGS - easy & flexible

winbugsfromstata - data preparation, analysis of MCMC output, graphics

Prior distributions - controversial

Check complex Stata models - vague prior distributions

Fit complex models not possible in Stata
Bayesian residuals and model checking [Lu et al., 2007]
Developments

- Bayesian residuals and model checking [Lu et al., 2007]
- Automate WinBUGS model statement
Developments

- Bayesian residuals and model checking [Lu et al., 2007]
- Automate WinBUGS model statement
- Mac users: WinBUGS runs under Darwine
Developments

- Bayesian residuals and model checking [Lu et al., 2007]
- Automate WinBUGS model statement
- Mac users: WinBUGS runs under Darwine
- OpenBUGS (version 3.0.1), WinBUGS (version 1.4.2)
  http://mathstat.helsinki.fi/openbugs/
A multilevel analysis of school examination results.

Meta-analysis of mixed treatment comparisons at multiple follow-up times.
*Statistics in Medicine.*
in press.

*WinBUGS User Manual, version 1.4.1.*
MRC Biostatistics Unit, Cambridge, UK.

Bayesian Analysis in Stata using WinBUGS.

Acknowledgements

MRC Capacity Building PhD Studentship in Genetic Epidemiology