

Using WinBUGS inside SAS for Bayesian Analysis

Zhiyong Zhang¹, John J. McArdle², Lijuan Wang¹,

& Fumiaki Hamagami¹

¹University of Virginia ²University of Southern California

Abstract

Bayesian methods are receiving more and more attention nowadays. To advance the application of Bayesian methods, we propose a standardized procedure to implement Bayesian analysis by interfacing WinBUGS and SAS, which allows use of WinBUGS inside SAS. This procedure is explained step by step by fitting a data set simulated from a linear growth curve model. Another example on how to iteratively run WinBUGS for Monte Carlo simulation studies is also provided.

Introduction

➤ Why use Bayesian methods?

✓ **Have been successfully applied to item response models (e.g., Chang, 1996; Fox & Glas, 2001), factor analysis models (e.g., Bartholomew, 1981; Lee, 1981), structural equation models (e.g., Scheines, Hoijsink, & Boomsma, 1999), genetic models (e.g., Eaves & Erkanli, 2003), and multilevel models (e.g., Seltzer, Wong, & Bryk, 1996).**

✓ **Rupp, Dey, and Zumbo (2004): “Both applied and theoretical communities cannot afford to miss the opportunities opened up by Bayesian methods.”**

➤ **Why use WinBUGS?**

- ✓ **Convenient for both simple and complex models**
- ✓ **Easy to program different models**
- ✓ **Supportive user groups**

➤ **Why use WinBUGS inside SAS?**

- ✓ **Benefit researchers who are familiar with SAS**
- ✓ **Provide a seamless way to use SAS data for Bayesian modeling**
- ✓ **Make use of WinBUGS easier and reduce mistakes**
- ✓ **Especially useful and convenient for the repetition of similar work, such as Monte Carlo simulation studies.**

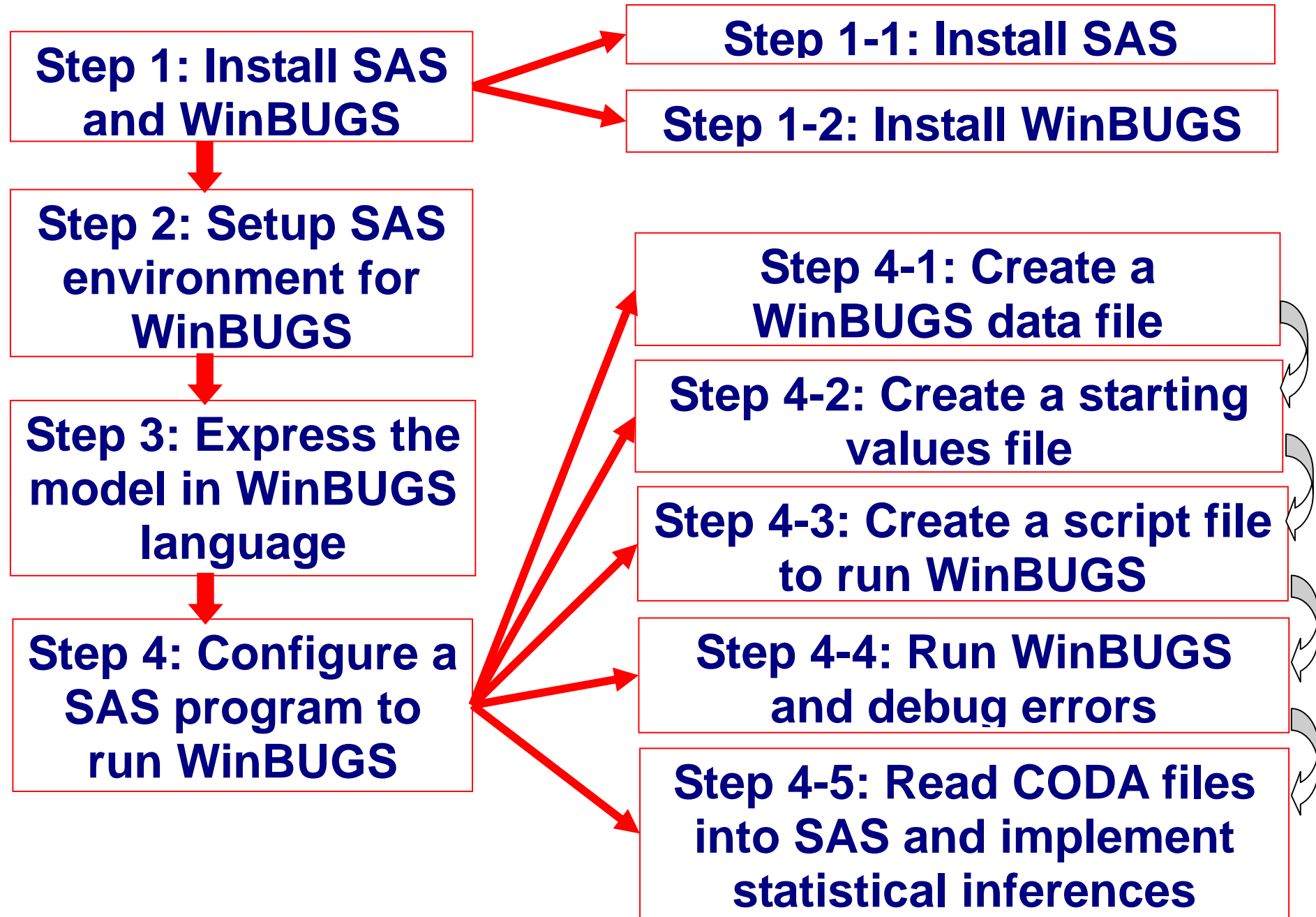
➤ **Two examples on how to implement Bayesian analysis inside SAS**

✓ **A linear growth curve model illustrates a step by step procedure to analyze one data set.**

✓ **A Monte Carlo simulation study on a confirmatory factor analysis model demonstrates how to iteratively run WinBUGS inside SAS.**

Corresponding should be addressed to Zhiyong Zhang, Department of Psychology, University of Virginia, Charlottesville, VA 22903. Email: zz5m@virginia.edu. A full version of this manuscript is available by request.

The flow chart to run WinBUGS inside SAS



Example 1. The Linear Growth Curve Model

- **One form of the linear growth curve model (McArdle & Nesselroade, 2003) is**

$$y_{it} = L_i + t \times S_i + e_{it}$$

$$L_i = \mu_L + v_{Li} \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

$$S_i = \mu_S + v_{Si}$$

- **Its probability form is**

$$y_{it} \sim N(\mu_{it}, \sigma_e^2)$$

$$\mu_{it} = L_i + t \times S_i \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

$$\begin{pmatrix} L_i \\ S_i \end{pmatrix} \sim MN \left(\begin{bmatrix} \mu_L \\ \mu_S \end{bmatrix}, \begin{bmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{bmatrix} \right)$$

Method

- A data set with $N=1000$ participants, and $T=5$ occasions was simulated with the true parameter values $\mu_L=10$, $\mu_S=5$, $\sigma_e^2=1$, $\sigma_L^2=4$, $\sigma_S^2=1$, and $\sigma_{LS}=1$ or $\rho_{LS}=.5$.
- The Bayesian estimation was implemented using WinBUGS inside SAS following the procedure proposed above. The complete SAS program is available by request zz5m@virginia.edu.
- The maximum likelihood estimation (MLE) was implemented using SAS PROC MIXED.

Results

➤ *Parameter estimates for the linear growth model*

	True Value	WinBUGS		SAS MIXED	
		Estimate	s.e.	Estimate	s.e.
μ_L	10	10.12	0.071	10.12	0.071
μ_S	5	5.01	0.033	5.01	0.034
σ_L^2	4	3.96	0.227	3.96	0.226
σ_S^2	1	1.04	0.051	1.04	0.051
σ_e^2	1	0.97	0.025	0.97	0.025
ρ	0.5	0.45	0.034	0.44	0.038

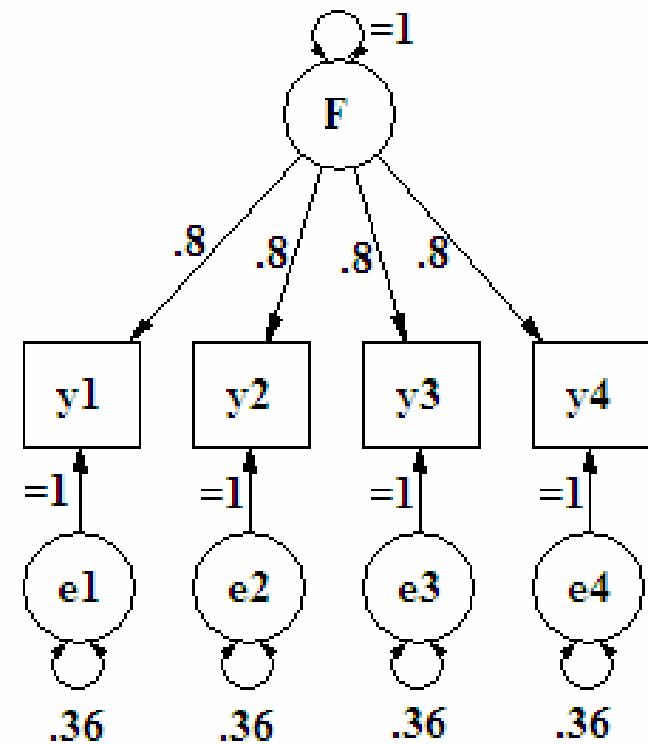
➤ *Bayesian method can obtain the same results as MLE when non-informative priors are used.*

Example 2. Monte Carlo Simulation Study of a Confirmatory Factor Model

- **Bayesian methods and WinBUGS have been used to estimate new models which usually cannot be done in the MLE framework yet.**
- **The evaluation of new estimation methods or new models through Monte Carlo studies is desired.**
- **WinBUGS itself is not very flexible for this purpose.**
- **This example demonstrates how to iteratively run WinBUGS inside SAS for Monte Carlo simulation and can be viewed as a supplement to WinBUGS.**

Method

- Data were simulated based on a confirmatory factor model.
- The path diagram for the true model is given as below with true parameter values on the paths.
- R=100 replications with the sample size N=200 were generated.
- Aim to show the Monte Carlo simulation procedure for using WinBUGS inside SAS.



Results

➤ *Parameter estimates for the CFM*

Parameters	TRUE	Mean	s.d.	MSE
	0.8	0.81	0.06	0.06
Factor	0.8	0.82	0.06	0.06
Loadings	0.8	0.80	0.06	0.06
	0.8	0.81	0.06	0.06
	0.36	0.35	0.04	0.05
Unique	0.36	0.37	0.05	0.05
Variances	0.36	0.38	0.05	0.05
	0.36	0.36	0.05	0.05

Note. Mean: the average value of the parameter estimates from 100 replications. s.d.: the standard deviation of the parameter estimates from 100 replications. MSE: the average standard errors of the parameter estimates from 100 replications.

- **In the real simulation study, one may consider more complex models and vary factor loadings and measurement errors.**
- **In terms of computing time,**
 - ✓ **For the growth curve model with $N=1000$ and $T=5$, it took a laptop (Celeron 1.7GHZ and 512M RAM) about 120 seconds.**
 - ✓ **For the confirmatory factor model, it took 30 minutes to finish the whole simulation study (with 100 replications).**

The complete SAS program to implement the analysis for CFM

CODE 1. The common part for the Monte Carlo simulation

```

FILENAME model "C:\SASWinBUGS\cfamodel.txt";
DATA model;
  INPUT model $80.;
  CARDS; /* model specification */
  model{
    for (i in 1:200){
      for (t in 1:4){
        y[i,t]~dnorm(muy[i,t],Inv_sig2[t])
        muy[i,t]<-fload[t]*fscore[i] }
      fscore[i]~dnorm(0, 1); escore[i]~dnorm(0,1)}
    for (t in 1:4){
      fload[t]~dnorm(0, 1.0E-6); Inv_sig2[t]~dgamma(0.001, .001)
      Para[t]<-fload[t]; Para[t+4]<-1/Inv_sig2[t]}}
  ; RUN;
DATA _NULL_; SET model; FILE model; PUT model; RUN;
/*Starting values*/
DATA _NULL_;
  FILE "C:\SASWinBUGS\cfaini.txt";
  PUT "list(fload=c(.5,.5,.5,.5), Inv_sig2=c(1,1,1,1))";
RUN;
/*Scripts to run WinBUGS*/
FILENAME runcfa 'c:\program files\winbugs14\runcfa.txt';
DATA _NULL_;
  FILE runcfa; PUT@1 "display('log')";

```

```

PUT@1 "check('C:/zzy/research/SASWinBUGS/cfamodel.txt') " ;
PUT@1 "data('C:/zzy/research/SASWinBUGS/cfadata.txt')";
PUT@1 "compile(1)";
PUT@1 "inits(1, 'C:/zzy/research/SASWinBUGS/cfaini.txt')";
PUT@1 "gen.inits()"; PUT@1 "update(2000)";
PUT@1 "set(Para)"; PUT@1 "update(3000)";
PUT@1 "stats(*)";
PUT@1 "save('C:/zzy/research/SASWinBUGS/cfalog.txt')";
PUT@1 "quit()"; RUN;
DATA _NULL_ ; /*create a .bat file to run WinBUGS*/
FILE "C:\SASWinBUGS\runcfa.bat";
PUT "C:\program files\WinBUGS14\WinBUGS14.exe" /PAR runcfa.txt';
PUT 'exit'; RUN;

```

CODE 2. The SAS macros to run the Monte Carlo simulation

```

%MACRO simcfa(n);
TITLE2 'Generate the Data';
DATA Sim_CFA;
*setting the true parameter values;
fload=.8; sig2=.36; N = 200; seed = 20060802+&n; M=4;
* need to setup arrays so we can have more variables;
ARRAY y_score{4} y1-y4; ARRAY e_score{4} y1-y4;
DO _N_ = 1 TO N;
    f_score=RANNOR(seed);
    DO t = 1 TO M;
        y_score{t} = fload*f_score +sqrt(sig2)*RANNOR(seed);
    END;
KEEP y1-y4; OUTPUT; END; RUN;

```

```
/*Transform Data to WinBUGS format*/
%_sexport(data=Sim_CFA, file='C:\SASWinBUGS\cfadata.txt',
var=y1-y4);
DATA _NULL_;/*Run WinBUGS in SAS X windows*/
  X "C:\SASWinBUGS\runcfa.bat"; RUN; QUIT;
/*Read in the log file to view parameter estimates*/
TITLE2 'Simulation '&n;
DATA log;
  INFILE "C:\SASWinBUGS\cfalog.txt" TRUNCOVER ;
  INPUT log $80.;
  log=translate(log," ","09"x);
  IF (SUBSTR(log, 2, 4) ne 'Para') then delete;
RUN;
PROC PRINT DATA=log; RUN;
%MEND;

%MACRO runsimcfa; /*Iteratively run WinBUGS*/
  %LET n=1;
  %DO %WHILE(&n <= 100);
    %simcfa(&n);
    %LET n=%EVAL(&n+1);
  %END;
%MEND runsimcfa;
*run the macro
%runsimcfa;
```

CODE 3.Analyze the results from the Monte Carlo simulation study

```

/*Save the output and log into files*/
DM OUTPUT 'FILE "C:\SASWinBUGS\allresults.txt"';
DM LOG    'FILE "C:\SASWinBUGS\allresults.log"';
/*Analyze the results*/
DATA temp;
  INFILE "C:\SASWinBUGS\allresults.txt" TRUNCOVER ;
  INPUT all $90.;
  IF (SUBSTR(all, 7, 4) NE 'Para') THEN DELETE;
  FILE "C:\SASWinBUGS\temp.txt";
  PUT all;
RUN;
DATA temp;
INFILE "C:\SASWinBUGS\temp.txt";
INPUT parid parname $ parest parsd MError p25 median p975 start sample;
id=int((_N_-.1)/8)+1; parest=abs(parest);
RUN;
/*Parameter Estimates*/
PROC TRANSPOSE DATA=temp OUT=parest PREFIX=par;
  BY id ;    ID parid;    VAR parest; RUN;
/*Calculate the mean and s.d. of the parameters*/
PROC MEANS DATA=parest; VAR par1-par8; RUN;
/*MSEs*/
PROC TRANSPOSE DATA=temp OUT=parsd PREFIX=sd;
  BY id ;    ID parid;    VAR parsd; RUN;

PROC MEANS DATA=parsd; VAR sd1-sd8; RUN;

```