The University of Iowa
ITP Iowa Testing Programs

The new Markov chain Monte Carlo (MCMC) procedure introduced in SAS/STAT® 9.2 and further exploited in SAS/STAT® 9.3 enables Bayesian computations efficiently with SAS®. The PROC MCMC procedure allows one to carry out complex statistical modeling within Bayesian frameworks under a wide-spectrum of scientific research; in psychometrics, for example, the estimation of item and ability parameters is a kind. This paper describes how to use the PROC MCMC procedures for Bayesian inferences of item and ability parameters under unidimensional item response models. How the results from SAS PROC MCMC procedures are different from or similar to the results from WinBUGS are also presented. For those who are interested in the Bayesian approach to item response modeling, it might be a good attempt to shift to SAS®, based on its flexibility of data managements and its power of data analysis.

PURPOSE OF THE STUDY

The study aims to compare and contrast the IRT parameter estimations from PROC MCMC procedure and WinBUGS, using both real and simulated data.

ITEM RESPONSE THEORY for DICHOTOMOUSLY SCORED ITEMS

Item response theory (IRT) uses a family of statistical models for estimating stable characteristics of items and examinees and defining how these characteristics interact in describing item and test performance. The three most famous IRT models for dichotomous response data are the one-, two-, and three-parameter logistic models by Rasch (1960), Birnbaum (1968), and Birnbaum (1968) and Lord (1980), respectively. The most generic form is as follows:

$$P_{ij}(\theta_j|a_i, b_i, c_i) = c_i + \frac{(1-c_i)}{1+\exp[-Da_i(\theta_j - b_i)]}.$$

- ❖Left side: the probability that the examinee j with an ability of θ_j answers the item i correctly
- ❖a_i: item discrimination of item i
- ❖b_i: item difficulty of item i
- ❖c_i: pseudo-chance of item I
- ❖D = 1.702; the scaling constant
- ❖2PL model: letting $c_i = 0$ for all items
- ❖1PL model: letting $c_i = 0$ and $a_i = 1$ for all items

ESTIMATION of PARAMETERS

Typically, the marginal maximum likelihood estimation along with an expectation-marginalization (Bock & Aitkin, 1981) and the marginal maximum a posteriori estimation (Mislevy, 1986) are used for IRT parameter estimation. Nowadays, the powerful simulation methods and the development of computing have made the Bayesian modeling possible in various research fields. The family of the MCMC methods is a popular way to simulate from a general posterior distribution which can be used for making inferences about model parameters.

PROC MCMC | WinBUGS

Resemblance

- ❖ Both are Markov chain Monte Carlo (MCMC) simulation procedures which are designed to fit Bayesian models or to make inferences of model parameters.
- ❖ Both provide sample autocorrelations and Monte Carlo errors for assessing convergence.
- ❖ In this study, the following remains the same across two estimation procedures: data being analyzed, priors of parameters of interest, initial values of the single chain, and numbers of MCMC iterations and burn-ins.

Main Difference

- ❖ Metropolis algorithm is for constructing the Markov chain. The Geweke , Raftery-Lewis and Heidelberger-Welch tests, and effective sample sizes are also available for assessing convergence.
- ❖ Gibbs sampling algorithm is for constructing the Markov chain. The Brooks, Gelman, and Rubin diagnostic is available for assessing convergence.

1 PL Model —Real Data

`nmc=100000`
`nbi=10000`

Number of parameters of interest: 310

PROC MCMC:

Item	$\hat{b}$ (scaled)	SD	MC error
1	-4.9002	0.8905	0.0539
2	-0.6572	0.1173	0.0041
3	-0.3983	0.1163	0.0035
4	-1.0817	0.1204	0.0037
5	-0.3726	0.1165	0.0038
6	-0.9770	0.1230	0.0037
7	-1.1290	0.1228	0.0034
8	-1.0431	0.1235	0.0031
9	0.5683	0.1196	0.0033
10	-1.1784	0.1239	0.0036

WinBUGS:

Item	$\hat{b}$ (scaled)	SD	MC error
1	-6.4667	0.2275	0.0008
2	-1.1564	0.0813	0.0003
3	-0.6461	0.0789	0.0003
4	-2.0017	0.0886	0.0003
5	-0.5997	0.0789	0.0003
6	-1.7977	0.0865	0.0003
7	-2.0957	0.0895	0.0003
8	-1.9127	0.0877	0.0004
9	1.2044	0.0816	0.0003
10	-2.1897	0.0905	0.0004

2 PL Model —Real Data

`nmc=300000`
`nbi=30000`

Number of parameters of interest: 320

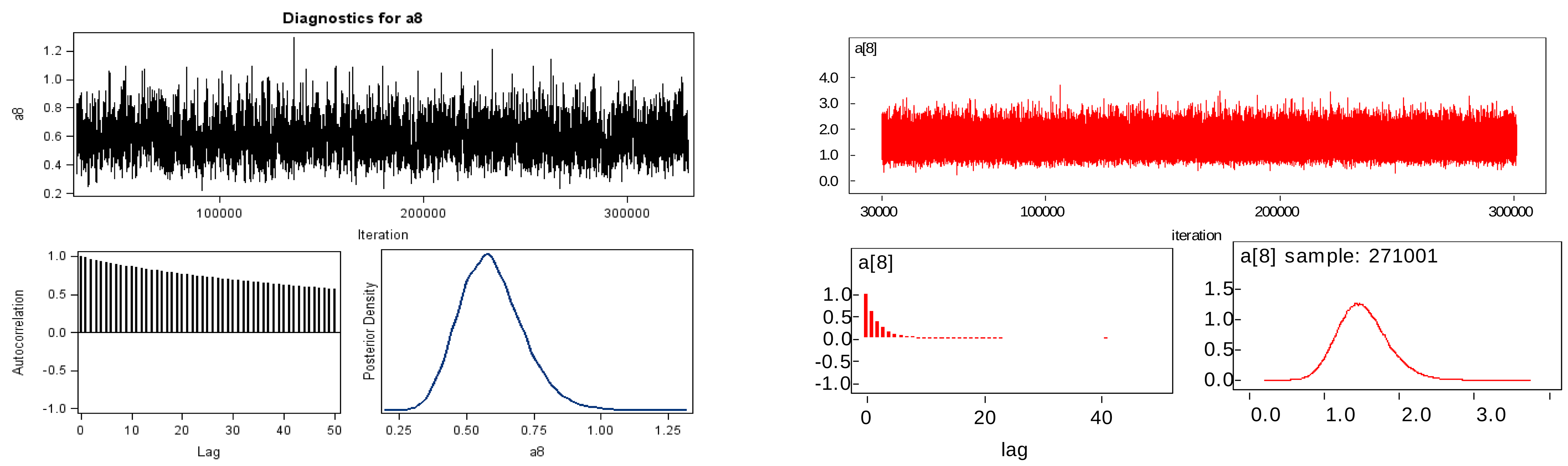

Figure 1 . Trace plots, autocorrelation plots, and kernel density of the posterior distributions of the discrimination parameter—Item 8 as an example (left panel—PROC MCMC; right panel—WinBUGS)

3 PL Model —Simulated

`nmc=600000`
`nbi=60000`

Number of parameters of interest: 330

PROC MCMC:

Item	$\hat{c}$	SD	MC error
1	0.2045	0.0617	0.0013
2	0.2388	0.0365	0.0006
3	0.2408	0.0873	0.0014
4	0.1769	0.0576	0.0012
5	0.2714	0.0957	0.0026
6	0.2523	0.0666	0.0016
7	0.1953	0.0668	0.0012
8	0.2353	0.0875	0.0021
9	0.2106	0.0374	0.0007
10	0.1631	0.0475	0.0008

WinBUGS:

Item	$\hat{c}$	SD	MC error
1	0.1938	0.0624	0.0003
2	0.2334	0.0372	0.0001
3	0.2411	0.0870	0.0004
4	0.1690	0.0549	0.0003
5	0.2581	0.0920	0.0005
6	0.2502	0.0664	0.0004
7	0.1971	0.0670	0.0003
8	0.2379	0.0860	0.0004
9	0.2000	0.0409	0.0001
10	0.1597	0.0468	0.0002

DISCUSSION AND CONCLUSIONS

- After scaling, the 1PL difficulty parameter estimates from PROC MCMC and WinBUGS are different but the correlation between the two sets is high ($r = 0.98$). Though not presented, for the 2PL and 3PL models, the correlations of the discrimination, difficulty, and pseudo-chance parameter estimates are also high (ranging from $r = 0.97$ to $r = 0.99$). For different models, the correlations of the ability estimates from PROC MCMC and WinBUGS are around 0.85.
- Across the board, *SD*s and MC errors are smaller for the estimates from WinBUGS.
- To achieve desirable accuracy, the number of iterations and the number of burn-ins are increased for 2PL and 3PL models. However, the autocorrelations are high from PROC MCMC.
- The estimation results for the pseudo-chance parameters are quite similar.
- With the most updated version, the PROC MCMC procedure requires relatively more time to obtain the posterior distributions.
- Limitation: the goodness of fit is not assessed for either 1PL or 2PL models.
- Future work: Using BILOG-MG as the anchor estimation program, it might be of interest to compare and contrast the estimation accuracy and results from the PROC MCMC procedure and WinBUGS/OpenBUGS against BILOG-MG.

References

- Birnbaum, A. (1968). Estimation of an ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 423–479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.
- Chen, F. (2009). Bayesian modeling using the MCMC procedure. *Proceedings of the SAS® Global Forum 2009 Conference*, Paper 257. Cary, NC: SAS Institute.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum Associates.
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, *51*, 177–195.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagogiske Institute.
- SAS Institute Inc. (2009). *Base SAS® 9.2 Procedures Guide*. Cary, NC: author.
- SAS Institute Inc. (2010). The MCMC procedure. In *SAS/STAT(R) 9.2 user's guide* (2nd Ed.) (author). Retrieved from http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#mcmc_toc.htm
- SAS Institute Inc. (2011). The MCMC procedure. In *SAS/STAT® 9.3 user's guide* (author). Retrieved from http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mcmc_sect002.htm
- Spiegelhalter, D. J., Thomas, A., & Best, N. G. (1999). WinBUGS version 1.4.3 [Computer program]. Available: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). *WinBUGS user manual*. Retrieved from http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf
- Thissen, D., Bock, R. D., & Muraki, E. (2003). Estimation. In M. du Toit (Ed.), *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT* (pp. 592–617). Lincolnwood, IL: Scientific Software International, Inc.

Acknowledgements

The author would like to thank Dr. Stephen Dunbar, the Iowa Testing Programs at the University of Iowa, for the support to this study. The author also thanked Dr. Hueying Tzou, National University of Tainan, for offering a sample of students' response data.

Contact the Author

Your comments and questions are valued and appreciated. Please contact the author at:

Yi-Fang Wu
Educational Measurement and Statistics, University of Iowa
E-mail: wuyifang91@gmail.com or yi-fang-wu@uiowa.edu

Sample code for the published examples as well as full discussion of the results are available from the author upon request.

SAS GLOBAL FORUM

THE UNIVERSITY OF IOWA

ITP Iowa Testing Programs