A Recent History of Bayesian Statistical Software

Rodney A. Sparapani, MS mailto:rsparapa@mcw.edu
Center for Patient Care and Outcomes Research
Medical College of Wisconsin

Prof. Purushottam W. (Prakash) Laud, PhD mailto:laud@mcw.edu
Division of Biostatistics
Medical College of Wisconsin

March 5, 2008

> You will find URLs for all of the software packages discussed here, and more, on the MILWASA web page for the Bayesian workshop at
> http://www.amstat.org/chapters/milwaukee/workshop

This history will, necessarily, consist of the opinions of the authors. Please read it in that vein. When you are delving into Bayesian statistical software, or any software for that matter, it is helpful to keep in mind the following advice from a renowned computer scientist:

> If all else fails, read the instructions. - Donald Knuth

The era of modern Bayesian statistical computation can be said to begin with the paper by Gelfand and Smith[1]. The construction of general purpose computational software, however, begins with three seminal papers by Gilks et al.[2,3] and Spiegelhalter et al.[4]. These advances culminated in the release of the free software package BUGS[5,6] (Bayesian inference Using Gibbs Sampling). BUGS software had two components: a model specification language, and a command language that could be utilized either interactively by the command line or in batch via a script file. BUGS was available for many Unix platforms as well as Linux and MS-DOS. BUGS relied on other software like the free software R[7] to create input data file(s) and to analyze its output data files such as the R packages CODA[8] (Convergence Diagnosis and Output Analysis for MCMC) or BOA[9] (Bayesian Output Analysis). BUGS although still available, is no longer maintained. BUGS was succeeded by the free software package WinBUGS[10].

WinBUGS is only available for MS Windows and is based on the BlackBox Component Builder developed by Oberon microsystems, a component-based development environment for the programming language Component Pascal. The model specification language is largely the same as that of BUGS, and WinBUGS still relies on other software to create input data file(s) as before. Interactive use is handled by the GUI of WinBUGS. Batch processing is handled by a new WinBUGS

command language which is not the same as the BUGS command language. WinBUGS also provides its own convergence diagnostics via the Gelman-Rubin statistics[11, 12] while still allowing you to create output data files to analyze as in the past. The R package R2WinBUGS[13] is a work-in-progress that manages the whole process from R: submitting the data and model file to WinBUGS, batch processing the MCMC sampling in WinBUGS and returning the samples to R. Although WinBUGS is an MS Windows application, it is currently possible to run it on other x86 platforms, like Unix/Linux and Mac OS X, via Wine, a free software, open source implementation of the MS Windows API for X11/OpenGL (and R2WinBUGS can take advantage of Wine as well).

WinBUGS is considered to be stable, but it will be phased out in the future. Current development is based on OpenBUGS[14], an open source version of WinBUGS that runs on MS Windows, Linux and as an R package. Although, OpenBUGS is in its early stages, OpenBUGS for MS Windows is quite robust and where new WinBUGS features are appearing. OpenBUGS shares much with WinBUGS including most of what has been described above like convergence diagnostics, R2WinBUGS and Wine. One difference is that OpenBUGS does not share the WinBUGS command language for batch processing, but instead has its own command language which is also not the same as the original BUGS command language. An advantage of OpenBUGS is you don't have to register it annually, something that was a minor irritant with WinBUGS. From here on out, the phrase OpenBUGS will refer to the MS Windows version and all comments will apply equally well to WinBUGS.

SAS® started out as a statistical analysis software package at a time when there were few options. Over time, SAS also built in capabilities that would facilitate data operations such as capture, management and manipulation. And, it is in this unique framework that SAS has prospered as one of the few annual fee software packages: you don't buy SAS, you "rent" it. SAS combines two levels of data programming: a low-level called the DATASTEP and a high-level known as Procedures or PROCs. SAS also provides the user with the SAS Macro Language: a facility for creating reusable SAS scripts called macros that can also provide high-level Procedure-like functionality.

Until recently, SAS did not offer much in the way of Bayesian statistics. SAS provides two SAS macros[15], `bayestests` and `bayesintervals`, for multiple testing and simultaneous intervals from the posterior sample. Also, with SAS you can perform Bayesian Variance Component analysis[16]. And very recently SAS has made available three Bayesian-capable PROCs: GENMOD, PHREG and LIFEREG which will be included in the next release of SAS. These PROCs are available as an experimental download on the MS Windows platform with the names BGENMOD, BPHREG and BLIFEREG. In addition, the user's manual[17] contains a nice introduction to Bayesian statistics. The 40 pages of material, including 7 pages of references, is worth reading for all who are interested in Bayesian statistics, whether they plan on using SAS or not, and whether they are novices or more advanced.

Currently, many Bayesians use R or SAS for its powerful data manipulation, and OpenBUGS for the statistical analysis. In this manner, the modeling and inference flexibility of OpenBUGS can be combined with the data manipulation and graphical power of R or SAS to explore the Markov chain Monte Carlo samples obtained from OpenBUGS.

For those using SAS, this process is facilitated by the free software, open source SAS macro library called RASmacro[18, 19]. It is a library of middle-level SAS macros that are the building

blocks for high-level SAS macros. RASmacro provides two SAS macros, `_lexport` and `_sexport`, to create input data for OpenBUGS. `_lexport` takes a list of SAS dataset variables and creates an input data file of scalars referred to as a "list" data file. `_sexport` takes a list of SAS dataset variables and creates an input data file of vectors referred to as a "structure" data file. RASmacro also provides two SAS macros, `_decoda` and `_debugs`, to process OpenBUGS output files. `_decoda` creates a SAS dataset from the OpenBUGS text output files: the index file and chain file(s). `_debugs` generates posterior statistics and plots, histograms and trace files (`_decoda` will call `_debugs` if statistics and graphics are requested).

In the last 20 years, we have seen Bayesian statistical software emerge from humble beginnings to the powerful applications that we have today. We have no doubt that this trend will continue. In the near future, we will most likely see new theoretical advances, better software and faster hardware. The confluence of these forces will make Bayesian statistics even more widespread than it is today.

References

[1] A Gelfand and A Smith, *Sampling-based approaches to calculating marginal densities*, Journal of the Americal Statistical Association **85** (1990), 398–409.

[2] W Gilks and P Wild, *Adaptive rejection sampling for Gibbs sampling*, Applied Statistics **41** (1992), 337–348.

[3] W Gilks, N Best, and K Tan, *Adaptive rejection Metropolis sampling with Gibbs sampling*, Applied Statistics **44** (1995), 455–472.

[4] D Spiegelhalter, A Thomas, and N Best, *Computation on Bayesian graphical models*, In Bayesian Statistics 5 (1995).

[5] D Spiegelhalter, A Thomas, N Best, and W Gilks, *BUGS: Bayesian inference using Gibbs sampling, Version 0.50*, MRC Biostatistics Unit, Cambridge, UK, 1995.

[6] N Best, D Spiegelhalter, A Thomas, and C Brayne, *Bayesian analysis of realistically complex models*, Journal of the Royal Statistical Society: Series A **159** (1996), 323–342.

[7] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2007, `http://www.R-project.org`. ISBN 3-900051-07-0.

[8] M Plummer, N Best, K Cowles, and K Vines, *CODA: Convergence Diagnosis and Output Analysis for MCMC*, March, 2006, `http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf`.

[9] B Smith, *Bayesian Output Analysis Program (BOA), Version 1.1.5*, The University of Iowa, March 23, 2005, `http://www.public-health.uiowa.edu/boa`.

[10] D Lunn, A Thomas, N Best, and D Spiegelhalter, *WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility*, Statistics and Computing **10** (2000), 325–337.

[11] Gelman A and Rubin D, *Inference from iterative simulation using multiple sequences*, Statistical Science **7** (1992), 457–472.

[12] Brooks S and Gelman A, *Alternative methods for monitoring convergence of iterative simulations*, Journal of Computational and Graphical Statistics **7** (1998), 434–455.

[13] S Sturtz, U Ligges, and A Gelman, *R2WinBUGS: A Package for Running WinBUGS from R*, 2005.

[14] A Thomas, B O'Hara, U Ligges, and S Sturtz, *Making BUGS Open*, March, 2006, `http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf`.

[15] Westfall, Tobias, Rom, Wolfinger, and Hochberg, *Multiple Comparisons and Multiple Tests Using the SAS® System*, SAS Institute Inc., Cary, NC, USA, 1999. ISBN 1-58025-397-0.

[16] Wolfinger R and Kass R, *Nonconjugate Bayesian analysis of variance component models*, Biometrics **56** (2000), 768–774.

[17] SAS Institute Inc. 2006, *Preliminary Capabilitites for Bayesian Analysis in SAS/STAT Software*, SAS Institute Inc., Cary, NC, USA, 2006.

[18] R Sparapani, *Some SAS for BUGS Data*, 2004, `http://ww2.mcw.edu/pcor/bugs/tr049.pdf`.

[19] ———, *RASmacro*, 2007, `http://ww2.mcw.edu/pcor/bugs/rasmacro.zip`.