

THE ISBA BULLETIN

Vol. 11 No. 2

June 2004

The official bulletin of the International Society for Bayesian Analysis

A MESSAGE FROM THE NEW PRESIDENT

by Jim Berger
ISBA President

berger@stat.duke.edu

Words here

A MESSAGE FROM THE EDITOR

by Hedibert Lopes
hedibert.lopes@gsb.uchicago.edu

Words here

Contents

- ▶ Applications
☛ Page 3
- ▶ Annotated Bibliography
☛ Page 6
- ▶ Bayesian History
☛ Page 11
- ▶ Software review
☛ Page 11
- ▶ News from the world
☛ Page 14

SUGGESTIONS

PLEASE, FEEL COMPLETELY FREE TO SEND US SUGGESTIONS THAT MIGHT
IMPROVE THE QUALITY OF THE BULLETIN

hedibert.lopes@gsb.uchicago.edu

FORMALIZING PROBABILITIES OF CLIMATE CHANGE THROUGH A BAYESIAN APPROACH

by Claudia Tebaldi
tebaldi@ucar.edu

The science of climate change is at a stage of undiscussed advancement, with complex numerical models (GCMs) able to run under alternative scenarios of future anthropogenic emissions, thus delivering projections of climate change at regional scales, and observational studies applying sophisticated statistical analyses in order to detect trends or shifts in the available records attributable to human impacts on climate rather than natural variability.

Climate policy discussions, with regard to both mitigation and adaptation issues are also well underway, and arguably sharpening, but they seem to be hampered by, and thus to denounce the lack of, a robust probabilistic evaluation of the future outcomes (Pittock et al. 2001, Reilly et al. 2001, Schneider 2001 and 2002, Webster et al. 2003).

Running GCMs for projections of climate change requires many choices of model parameters and future scenarios, so an evaluation of these inputs' relative probability is unavoidable (Soon et al. 2001, Visser et al. 2000, Stott and Kettleborough 2002). It is either a matter of attaching degrees of belief in alternative hypothesis of future emissions, or of defining likely ranges of variations for numerical models' parameters, or of formalizing a priori the reliance on a particular GCM's (regional) projections over another. In any case, strength for the ensuing evaluation of the results of climate model experiments can be gained by formalizing this information through the elicitation of experts' opinion and their combination with observational records and numerical model output (Carter et al. 2001, Nordhaus 1994, Risbey and Kardlikar 2002, Raper et al. 2001, Wigley and Raper 2001, Webster and Sokolov 2000).

The Bayesian approach is naturally suited to handling these issues and has been in fact recognized as such by many members of the climate research community, so much so that coming across some sort of "Bayesian" analysis while browsing through recent articles in the climate research literature is more and more frequent. Model selection, meta analysis, analysis of computer experiments, hierarchical modeling and MCMC solutions all have some bearing over the formalization of probabilistic forecasts of climate change.

As a statistician of Bayesian upbringing working at the National Center for Atmospheric Research

(NCAR, see <http://www.ncar.ucar.edu>), in collaboration with Doug Nychka and the other members of his Geophysical Statistics Project (GSP, see <http://www.cgd.ucar.edu/stat>) with the added advantage of regular visits by Richard W. Smith to our center, I have the opportunity of tackling some of these issues first hand.

Together with Richard, Doug, and Linda O. Mearns, a physical geographer, scientist in the Environmental and Societal Impacts Group (ESIG, see <http://www.esig.ucar.edu>) at NCAR, whose research has been geared for many years towards translating the scientific findings of climate studies into meaningful societal impacts, we have formulated a Bayesian model for combining projections of temperature and precipitation changes for regional domains (of subcontinental size) from nine different GCMs, run under the same scenarios. Model uncertainty is an important component of the uncertainty in these projections, and different GCMs disagree significantly in their projections, even under the same forcing scenarios. Thus, the treatment of an ensemble of GCMs is a necessary step in trying to quantify between-model variability and separate it from within model uncertainty. In the simplest version of our statistical model, we treat present and future mean climate and each GCM's accuracy as stochastic parameters. We are building a Bayesian version of a random effect model in which GCMs are thought as...rats: a limited sample from a superpopulation of all possible (present and future) GCMs, all aiming at the "true" climate signal, each of them with a model-specific degree of accuracy. Our data consists of the GCMs' responses for present and future climate averages, and present day observations of climate. The latter are to be compared to the corresponding model output, in order to derive a measure of model bias. The Bayesian framework naturally allows the formalization of all the sources of uncertainty in the analysis, both model- and physics-related. At the end of the day (or rather of the MCMC simulation), the posterior estimates of climate change have the intuitive characteristics of a weighted mean of the nine GCMs' projections, with weights that reward GCMs accurate in reproducing current climate (conversely downweighting biased GCMs) and GCMs agreeing with one another (conversely downweighting outliers in the ensemble). The heuristic criteria of model bias and inter-model agreement have been used in climate research studies in an ad-hoc fashion (Giorgi and Mearns 2002 and 2003). Our result is to formalize them as a consequence of a specified statistical model whose assumptions are easily determined and possibly crit-

icized.

Another project underway, this one directly addressing GCMs' parameterizations as a source of uncertainty, is in collaboration with Chris E. Forest of MIT and Bruno Sanso of UCSC. The goal is a formal Bayesian analysis of the probability distributions of three key parameters of MIT's intermediate GCM, related to the warming response of the model under a standard doubling of CO₂, to the negative forcing from aerosols in the atmosphere and to the oceans' capacity of heat uptake. The values of these three parameters affect the magnitude of the projected change in climate, and their ranges may be evaluated through the merging of expert opinion and model validation studies (Forest et al. 2001 and 2002). Many experiments are run, under different settings of the three parameters, and aimed at recreating present day climate. A statistical approximator to the GCM output is fitted, using thin plate splines, in order to obtain an efficient mapping from the three-parameter space to the model diagnostics that we are going to compare to observed records. Once the approximator is estimated, we can implement a full-fledged MCMC simulation of the joint posterior distribution of the three parameters, within which the approximator serves as a necessary step in the computation of the likelihood. A straightforward nesting of the MIT model within the MCMC iteration scheme is unfeasible because of the computing time associated with each of its runs. The output of the analysis may guide future simulations of the MIT model towards the use of particularly "likely" settings, that will in turn produce the most likely projections of climate change.

These are just two examples of the applications of Bayesian modeling to relevant questions in the climate change debate. This field is rich of interesting scientific questions, and challenging data sets, and offers many occasions for valuable (and fun) applications of Bayesian analysis. The results, in the form of probability distributions, are naturally appropriate input for many pressing policy questions.

So...tell your friends and neighbors.

References

- Carter, T.R., E. L. La Rovere, R. N. Jones, R. Leemas, L.O. Mearns, N. Nakicenovic, A. B. Pittock, S. M. Semenov and J. Skea (2001) Developing and applying scenarios, 145-190. In IPCC (ed.) *Climate change 2001: impacts, adaptation and vulnerability*, Cambridge University Press, Cambridge.
- Forest, C. E., M. R. Allen, A. P. Sokolov and P.

- H. Stone (2001) Constraining climate model properties using optimal fingerprint detection methods. *Climate Dynamics*, **18**, 277-295.
- Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen and M. D. Webster (2002) Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113-117.
- Giorgi F. and L.O. Mearns (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *Journal of Climate* **15**, 1141-1158.
- Giorgi F. and L.O. Mearns (2003) Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. *Geophysical Research Letters* **30**, art. no. 1629.
- Nordhaus, W. D. (1994) Expert opinion on climatic change. *American Scientist* **82**, 45-51.
- Pittock, A. B., R. N. Jones and C.D. Mitchell (2001) Probabilities will help us plan for climate change. *Nature* **413**, 249.
- Raper, S.C.B., J. M. Gregory and T. J. Osborn (2001) Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Climate Dynamics* **17**, 601-613.
- Reilly, J., P. H. Stone, C. E. Forest, M.D. Webster, H.D. Jacoby and R.G. Prinn (2001) Climate change- Uncertainty and climate change assessment. *Science* **293**, 430-433.
- Risbey, J. and M. Kandlikar (2002) Expert assessment of uncertainties in detection and attribution of climate change. *Bulletin of the American Meteorological Society* **83**, 1317-1326.
- Schneider, S.H. (2001) What is 'dangerous' climate change? *Nature* **411**, 17-19.
- Schneider, S.H. (2002) Can we estimate the likelihood of climatic changes at 2100? *Climatic Change* **52**, 441-451.
- Soon, W., S. Baliunas, S. B. Idso, K. Y. Kondratyev and E. S. Posmentier (2001) Modeling climatic effects of anthropogenic carbon dioxide emissions: unknown and uncertainties. *Climate Research* **18**, 259-275.
- Stott, P.A. and J.A. Kettleborough (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416**, 723-726.
- Visser, H., R.J.M. Folkert, J. Hoekstra and J.J. De Wolff (2000) Identifying key sources of uncertainty in climate change projections. *Climatic Change* **45**, 421-457.
- Webster, M.D., C. E. Forest, J. Reilly, M. Babiker, D. Kicklighter, M. Mayer, R. Prinn, M. Sarofim, A. P. Sokolov, P. Stone and C. Wang (2003) Un-

certainty analysis of climate change and policy responses. *Climatic Change* (submitted).
Webster, M.D. and A. P. Sokolov (2000) A methodology for quantifying uncertainty in climate projections. *Climatic Change* **46**, 417-

446.
Wigley, T. M. L. and S.C.B. Raper (2001) Interpretation of high projections for global-mean warming. *Science* **293**, 451-454.

ISBA/SBSS ARCHIVE FOR ABSTRACTS

All authors of statistics papers and speakers giving conference presentations with substantial Bayesian content should consider submitting an abstract of the paper or talk to the ISBA/SBSS Bayesian Abstract Archive. Links to e-prints are encouraged. To submit an abstract, or to search existing abstracts by author, title, or keywords, follow the instructions at the abstract's web site,
www.isds.duke.edu/isba-sbss/

ISBA Bulletin, June 2004

ANNOTATED BIBLIOGRAPHY

EVALUATION OF REPRODUCIBILITY

by Mosuk Chow
mchow@stat.psu.edu

Evaluation of reproducibility is needed for many scientific research problems. For example, when a new instrument is developed, it is of interest to assess whether the new instrument can reproduce the results obtained by using a traditional gold standard criterion. Indeed, the need to quantify agreement arises in many research fields when two approaches or two raters simultaneously evaluate a response. There are lots of literature on the subject but not much of them use a Bayesian approach.

Barnhart, H. X. and Williamson, J. M. (2001). Modeling concordance correlation via GEE to evaluate reproducibility. *Biometrics*, **57**, 931-940.

In this article, the authors propose a generalized estimating equations (GEE) approach to model the concordance correlation coefficient via three sets of estimating equations. The proposed approach is flexible in that (1) it can accommodate more than two correlated readings and test for the equality of dependent concordant correlation estimates; (2) it can incorporate covariates predictive of the marginal distribution; (3) it can be used to identify covariates predictive of concordance correlation; and (4) it requires minimal distribution assumptions.

Chinchilli, V. M., Martel, J. K. Kumanyika, S., and Lloyd, T. (1996). A weighted concordance correla-

tion coefficient for repeated measurement designs. *Biometrics*, **52**, 341-353.

The authors develop weighted product-moment and concordance correlation coefficients which are applicable for repeated measurements study designs. They consider two distinct situations in which the repeated measurements are paired or unpaired over time. They illustrate the methodology with examples comparing (1) two assays for measuring serum cholesterol, (2) two estimates of dietary intake, from a food frequency questionnaire and dietary recalls, and (3) two measurements of percentage body fat, from skinfold calipers and dual energy x-ray absorptiometry.

King, T. S and Chinchilli, V. M. (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Statistics in Medicine*, **20**, 2131-2147.

This paper discusses a generalized version of the concordance correlation coefficient for agreement data. The concordance correlation coefficient evaluates the accuracy and precision between two measures, and is based on the expected value of the squared function of distance. The authors have generalized this coefficient by applying alternative functions of distance to produce more robust versions of the concordance correlation coefficient. In this paper they extend the application of this class of estimators to categorical data as well, and demonstrate similarities to the kappa and weighted kappa statistics

Li, R. and Chow, M. (2004). Evaluation of reproducibility for paired functional data. *Journal of Mul-*

tivariate Analysis, in press.

Evaluation of reproducibility is important in assessing whether a new method or instrument can reproduce the results from a traditional gold standard approach. In this paper, the authors propose a measure to assess measurement agreement for functional data which are frequently encountered in medical research and many other research fields. Formulae to compute the standard error of the proposed estimator and confidence intervals for the proposed measure are derived.

Lin, L. I. (1989). A concordance correlation coefficients to evaluate reproducibility. *Biometrics*, **45**, 255-268.

In this paper, the author proposes a new reproducibility index. This index is the correlation between the two readings that fall on the 45 degree line through the origin. It is simple to use and possesses desirable properties.

Lin, L., Hedayat, A. S., Sinha, B. and Yang, M. (2002). Statistical methods in assessing agreement: models, issues and tools, *Journal of the American Statistical Association*, **97**, 257-270.

In this paper, the authors study various measurements of agreement needed to assess the ac-

ceptability of a new or generic process, methodology, and formulation in areas of laboratory performance, etc. In all of these areas, one needs measurements that capture a large proportion of data that are within a meaningful boundary from target values. Target values can be considered random (measured with error) or fixed (known), depending on the situation.

Quan, H. and Shih, W. J. (1996). Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics*, **52**, 1195-1203.

In this paper the authors consider the use of within-subject coefficient of variation (WCV) for assessing the reproducibility or reliability of a measurement. Application to assessing reproducibility of biochemical markers for measuring bone turnover is described and the comparison with intraclass correlation is discussed. Both maximum likelihood and moment confidence intervals of WCV are obtained through their corresponding asymptotic distributions. Normal and log-normal cases are considered. In general, WCV is preferred when the measurement scale bears intrinsic meaning and is not subject to arbitrary shifting. The intraclass correlation may be preferred when a fixed population of subjects can be well identified.

HAROLD JEFFREYS AND R. A. FISHER

by John Aldrich

john.aldrich@soton.ac.uk

1 Introduction

This note traces the differences between Harold Jeffreys (1891-1989) and Ronald Fisher (1890-1962) over probability and statistical inference. Fisher was once the central figure in statistics but Jeffreys was only ever on the fringe. His statistical work, summarised in *Theory of Probability* (1939, -48, -60), seemed misconceived both to Fisher, who thought he had destroyed its Bayesian basis long before, and to younger statisticians for whom Fisher was himself a dinosaur.

Jeffreys and Fisher had much in common. They were applied mathematicians whose research, in physics and genetics respectively, combined mathematical theory and empirical analysis. Their education was similar. Fisher entered Cambridge University as a mathematics undergraduate in 1909

and Jeffreys followed a year later. As an undergraduate, Fisher read Karl Pearson on biometry and found connections with what he was learning about the combination of observations. Jeffreys also learnt about combining observations but he used the methods for many years before contributing to the theory. Jeffreys also read Pearson and adopted his view that laws are not established with certainty but can have a high degree of probability on the data. He developed and applied this view in philosophy of science first and then in statistics.

Jeffreys wrote about scientific method with Dorothy Wrinch (1894-1976), who had links to the Cambridge logicians, including Bertrand Russell and W. E. Johnson. Their first paper "On Some Aspects of the Theory of Probability" (1919) presented the logicians' notion of probability as a degree of reasonable belief to a scientific audience. The paper also attacked the limiting frequency notion of probability. Wrinch and Jeffreys investigated the problem of sampling from an urn to determine its composition, but only to get insight into the process of accumulating evidence for scientific laws.

Jeffreys's career was not built on philosophy but on his work on the solar system and the earth.

Empirical seismology gradually became his main field and this stimulated him to do research on combining observations. Fisher, meanwhile, was making a career at Rothamsted agricultural station. He maintained his old interests in biometry/genetics/evolution and in statistical theory while creating the analysis of variance and the design of experiments. His *Statistical Methods for Research Workers* (1925) became the statistics book of the era. Fisher's notion of probability was based on relative frequency and he rejected Bayesian arguments unless there was a prior based on frequency data. He identified the (unsound) Bayesian argument with the use of a uniform prior, noting (1922, p. 325) that the inference depends on the parametrisation chosen.

2 The exchange 1933-34

In 1933 Jeffreys and Fisher came in sight of one another and both fired. They knew enough of each other's work to know it was futile: Jeffreys had exploded Fisher's notion of probability and Fisher had refuted the Bayesian principle. Given this, there is nothing surprising in Lane's (1980, p. 159) reflection on the exchange, "Nothing was settled Neither scientist seems to have convinced the other of anything."

Jeffreys was looking for better ways of analysing seismological data but he was also revisiting the Wrinch-Jeffreys topics. Among the new material in his *Scientific Inference* (1931) was a treatment of the fundamental problem in the combination of observations: inference to the true value, x , when the errors are normally distributed. The known precision case, which had been treated by Gauss before he went "Gauss-Markov", was in the textbooks but the unknown precision case was not. Jeffreys thought a uniform prior appropriate for x but he saw "no special reason for measuring the precision in terms of h [the established measure in error theory] rather than $\sigma [= \frac{1}{h\sqrt{2}}]$." He recognised the problem of the non-invariance of inference to reparametrisation and disarmed it by making the prior proportional to $1/h$. This met Fisher's objection of 1922, though Jeffreys did not mention this. Nor did Fisher pay much attention to Jeffreys's use of invariant priors, even after Jeffreys developed the idea further in 1946.

In a paper extending the argument to regression, Jeffreys (1932) put the case for $1/h$ in an "alternative form": instead of asking, what distribution describes our prior knowledge, we ask what distribution is consistent with facts otherwise known about the posterior probability on certain types of data. Jeffreys (1932, p. 48) showed that only $1/h$ gener-

ates a predictive density embodying the following "fact":

Two measures are made: what is the probability that the third observation will lie between them? The answer is easily seen to be one-third. For the law says nothing about the order of occurrence of errors of different amounts, and therefore the middle one is equally likely to be the first, second, or third made (provided, of course, that we know nothing about the probable range of error already.)

Fisher (1933, p. 344) ridiculed the argument on the (expected) ground, "That there should be a method of evolving such a piece of information by mathematical reasoning only ... would be in all respects remarkable ..." but he also thought there was a mathematical mistake. The discussion is further tangled by Fisher's desire to describe his new (from 1930) fiducial argument and to distinguish it from Jeffreys's argument.

Simultaneously Jeffreys (1933, p. 86) was firing at Fisher: "The whole reason for attaching any importance to Fisher's 'likelihood' is that it is proportional to the posterior probability given by Laplace's theory, and it has no meaning outside the original sample except in terms of that theory." Fisher's criticism of his work on $1/h$ brought an immediate reply. Jeffreys (1933a) explained he was not trying to "evolve a piece of information" in Fisher's sense. Under fire he became bolder. *Scientific Inference* had not fully endorsed the $1/h$ prior for there was a modification to make the integral unity, to conform to the "convention" that "1 is the constant to be attached to certainty." Now, however, impropriety became a virtue and Jeffreys (1933a, p. 531) gave new arguments to support "the fact ... that my distribution is the only distribution of prior probability that is consistent with complete ignorance of the value of h ."

Jeffreys (1933a, p. 532) also made a general point about the averaging in Fisher's argument.

This procedure [of integrating with respect to all values of the observed measures] involves a fundamental confusion, which pervades the whole of his statistical work and deprives it of all meaning. The essential distinction in the problem of inference is the distinction between what we know and what we are trying to find out: between the data and the proposition on the data we are trying to find out. ... To integrate with respect to them and average

a function of them over the range of integration is an absolutely meaningless process.

This seems to be the first time Jeffreys noted in print the practice of averaging over the sample space. He returned to its meaninglessness when he criticised prob-values and conventional significance testing in the *Theory of Probability*.

The engagement (1934) ended with the protagonists producing a paper summarising his position. The parties separated with their authority intact and convictions reinforced. The only criticism that moved Jeffreys moved him to a more extreme position.

3 Reconciliation—of a kind

The encounter with Fisher *did* have an effect on Jeffreys. He changed his mind about *garbage in, garbage out* and studied statistical theory, re-doing it so it made sense. The outcome, *Theory of Probability* (1939), could have not been foreseen in 1931 or -33. Added to the philosophy, which went back to the collaboration with Wrinch, and to his work with data was the application of Bayesian methods to a range of problems comparable to that in Fisher's *Statistical Methods* with a Bayesian take on such Fisherian items as sufficiency, information and the fiducial argument.

This passage (p. 324) captures the book's tone towards Fisher:

I have in fact been struck repeatedly in my own work, after being led on general principles to a solution of a problem, to find that Fisher has already grasped the essentials by some brilliant piece of common sense, and that his results would be either identical with mine or would differ only in cases where we should both be very doubtful.

The book could have been subtitled *Fisher done properly*.

Fisher did not budge from *garbage in, garbage out* and there is little to suggest that he ever paid detailed attention to what Jeffreys wrote. Yet some of the points Jeffreys made against him reappeared when he criticised Neyman and his followers. Fisher's caricature of them as mathematicians with no experience of science could not be applied to Jeffreys.

In 1934 the journal brought Fisher and Jeffreys together to arrange the end of hostilities. Thereafter they enjoyed good personal relations. (Some of their letters are in Bennett (1990)). Fisher tended to

personalise intellectual differences but Jeffreys was a special case. There was no question of his betraying Fisher (and truth) because he *never* knew any better; nor was his work a threat because in Fisher's life-time very few statisticians took it seriously—it was an intellectual curiosity. When Fisher's fiducial argument came under attack from Bartlett (representing the next generation) and Neyman (a rival leader) Jeffreys's support was welcome. Yet Jeffreys only gave it because he believed that the fiducial argument *properly stated* was equivalent to his own argument. When the missing steps are supplied "the arguments are much longer than those got by introducing the prior probability to express previous ignorance at the start." (1939, p. 312)

4 Literature

This account is based on Aldrich (2003). Other works include a book on Jeffreys and Fisher by David Howie, a historian of science, and an article on the controversy by David Lane. For further information see my website *Harold Jeffreys as a Statistician*, which also has the urls of literature available on the web. For Fisher see my website *A Guide to R. A. Fisher*.

References

- Aldrich, J. (2003) The Statistical Education of Harold Jeffreys, Southampton University Economics Discussion Paper, <http://www.socsci.soton.ac.uk/Economics/Research/Discussion.Papers>
- _____ (2003a) *Harold Jeffreys as a Statistician*, <http://www.economics.soton.ac.uk/staff/aldrich/jeffreysweb.htm>
- _____ (2003b) *A Guide to R. A. Fisher*, <http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm>
- Bennett, J. H. (1990) (ed) *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*, Oxford, Oxford University Press.
- Fisher, R. A. (1922) On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society, A*, **222**, 309-368.
- _____ (1925) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.
- _____ (1933) The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters, *Proceedings of the Royal Society A*, **139**, 343-348.
- _____ (1934) Probability, Likelihood and the Quantity of Information in the Logic of Uncertain Inference, *Proceedings of the Royal Society, A*, **146**, 1-8.

Howie, D. (2002) *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*, New York, Cambridge University Press.

Jeffreys, H. (1931) *Scientific Inference*, Cambridge: Cambridge University Press.

_____ (1932) On the Theory of Errors and Least Squares, *Proceedings of the Royal Society, A*, **138**, 48-55.

_____ (1933) On the Prior Probability in the Theory of Sampling, *Proceedings of the Cambridge Philosophical Society*, **29**, 83-87.

_____ (1933a) Probability, Statistics and the Theory of Errors, *Proceedings of the Royal Society, A*, **140**,

523-535.

_____ (1934) Probability and Scientific Method, *Proceedings of the Royal Society, A*, **146**, 9-16.

_____ (1939) *Theory of Probability*, (2nd and 3rd editions in 1948 and -60), Oxford, University Press.

Lane, D. A. (1980) Fisher, Jeffreys and the Nature of Probability, pp. 148-160 in S. E. Fienberg & D. V. Hinkley (1980) (eds) *R., A. Fisher: An Appreciation* New York, Springer.

Wrinch, D. & H. Jeffreys (1919) On Some Aspects of the Theory of Probability, *Philosophical Magazine*, **38**, 715-731.

ISBA Bulletin, June 2004

SOFTWARE REVIEW

SOME SAS MACROS FOR BUGS/WINBUGS DATA

by Rodney Sparapani
rsparapa@mcw.edu

BUGS and WinBUGS are great tools for performing Bayesian analysis. However, their data manipulation capabilities are limited. Along with Matthew Hayat, I have developed some SAS macros that will allow you to create BUGS input data files from SAS datasets and to create SAS datasets from CODA output files. They are available on the internet at <http://www.mcw.edu/pcor/bugs/decoda.zip> with a brief discussion including some notes and documentation at <http://www.mcw.edu/pcor/bugs/>.

I have been working with SAS for over a decade and with SAS macros, in particular, for about ten years. SAS macros are very powerful, but often suffer from poor documentation. My own SAS macros are no exception. However, this is an opportunity to document my work and share it with others. I'm going to describe a set of SAS macros that were written to allow a seamless movement of data between SAS datasets to BUGS input files and CODA output files back to SAS datasets.

There are two SAS macros to make BUGS input files: `_LEXPOR`T to make "lists" (vectors) and `_SEXPOR`T to make BUGS-style "structures" (matrices). Here's a code snippet:

```
%_lexport(data=mta, file=list.txt,
          var=med beh com);
%_sexport(data=mta, file=struct.txt,
          var=pscale0-pscale3 tscale0-tscale3);
```

This creates one file of "list" input, `list.txt`, which contains three vectors (`med`, `beh` and `com`) and another file of BUGS-style "structure" input, `struct.txt`, which contains two matrices (`pscale` and `tscale`). Each of these SAS macros requires

only two parameters: `FILE=` and `VAR=`. `DATA=` is optional and it defaults to the last SAS dataset created. Other optional settings are `LS=` for line length and `FORMAT=` for variable length which you can change to make your input files more readable (e.g. for dummy variables, `LS=78`, `FORMAT=1.`).

After generating posterior samples with BUGS, we want to create SAS datasets from our CODA files. Here's a code snippet:

```
goptions device=jpeg gaccess=gsasfile
          gsfmode=replace;
%coda2sas(out=post1, infile=mta.ind,
          chain=mta1.out, stats=1);
goptions device=pslepsz;
%coda2sas(out=post2, infile=mta.ind,
          chain=mta2.out, stats=1, type=eps);
%_decoda(out=post3, infile=mta.ind,
          chains=2);
%bayesintervals(vars=c_1-c_3, tail=U);
```

The `CODA2SAS` macro was originally written with BUGS in mind. It can process either BUGS or WinBUGS CODA files, but doesn't handle multiple chains automatically. You have to import each of the chains manually. Set `INFILE=` to the name of your index file and `CHAIN=` to the name of your chain file. So, I imported the first chain into the SAS dataset `post1`. This SAS dataset contains three variables `c_1-c_3` which correspond to the monitored array `c[]` where `c[1]` was translated as `c_1`, etc. But first, I set my graphics device with a `GOPTIONS` statement because when you request a summary (`STATS=1`), statistics and kernel density plots with histograms are generated. Without a `GOPTIONS` statement like the one above, the plots will be displayed on your graphics device. However, if you choose a device that is a graphics file type (like `JPEG`), then graphics files are generated for each variable (`c_1.jpg`, etc.). If a different graphics file type is declared, then you must also set the `TYPE=` option which defaults to `jpg`. For example, the sec-

ond chain is imported into post2 and *encapsulated PostScript* graphics files are created (c_1.eps, etc.).

Next, I read in both chains with the `_DECODA` macro. Name your index file either `NAME.ind`, `NAMEIndex.txt` or `NAME.ind.txt` and each of your chains `NAME#.out`, `NAME#.txt` or `NAME#.out.txt` respectively, where # is the number of your chains (1-2 in this example). If you follow this naming convention, then an unlimited number of chains are supported by specifying only `INFILE=` for the index file and `CHAINS=` for the number of chains. If the chain files follow some other naming convention, then you can specify each of them manually as `CHAIN1=` up to `CHAIN10=`. The `INFILE=` and `OUT=` parameters are required as well as either `CHAINS=` or `CHAIN1=`, etc. The posterior samples from both chains are contained in the SAS dataset `post3`. The same `CODA2SAS` comments apply to `_DECODA` with respect to variable names, `STATS=` and `TYPE=`. One additional feature is that tests and tables for location are also performed and the default location of 0 can be changed with `MU0=`. Lastly, I used the SAS macro `BAYESINTERVALS` by RD Wolfinger which is available at <http://ftp.sas.com/samples/A56648>. It constructs simultaneous intervals of the posterior for `c_1-c_3`.

There are two other SAS macros that you might find useful: `SAS2CODA` and `_LIMPORT`. `SAS2CODA` with similar syntax to `CODA2SAS` (except there is no `STATS=` option) reads a SAS dataset and creates CODA files. `_LIMPORT` creates a SAS dataset from a "list" file with two required options, `OUT=` for the new SAS dataset and `INFILE=` for the "list" file. However, importing can be tricky so a temporary SAS/IML program (which you can name with `FILE=`) is created and run automatically. If the importing fails, then you should be able to make corrections to the program and run it manually.

`CODA2SAS`, `SAS2CODA` and `BAYESINTERVALS` are self-contained and do not require any other SAS macros. However, `_DECODA`, `_LEXPOR`, `_SEXPOR` and `_LIMPORT` require other SAS macros which are also included in `decoda.zip` (currently there are 18 others, but knowledge of them is not necessary). Here are the instructions to install all SAS macros contained in `decoda.zip` (similar instructions would apply to other SAS macros):

1. unzip `decoda.zip` in a directory I'll generically call `SASMACRO` (you can call it anything you want)
2. add the following line to your SAS configuration file (`sasv7.cfg`, `sasv8.cfg` or `sasv9.cfg`)
`-insert sasautos SASMACROBUGS and WinBUGS`

are great tools for performing Bayesian analysis. However, their data manipulation capabilities are limited. Along with Matthew Hayat, I have developed some SAS macros that will al-

low you to create BUGS input data files from SAS datasets and to create SAS datasets from CODA output files. They are available on the internet at <http://www.mcw.edu/pcor/bugs/decoda.zip> with a brief discussion including some notes and documentation at <http://www.mcw.edu/pcor/bugs/>.

I have been working with SAS for over a decade and with SAS macros, in particular, for about ten years. SAS macros are very powerful, but often suffer from poor documentation. My own SAS macros are no exception. However, this is an opportunity to document my work and share it with others. I'm going to describe a set of SAS macros that were written to allow a seamless movement of data between SAS datasets to BUGS input files and CODA output files back to SAS datasets.

There are two SAS macros to make BUGS input files: `_LEXPOR` to make "lists" (vectors) and `_SEXPOR` to make BUGS-style "structures" (matrices). Here's a code snippet:

```
%_lexport(data=mta, file=list.txt,
          var=med beh com);
%_sexport(data=mta, file=struct.txt,
          var=pscale0-pscale3 tscale0-tscales3);
```

This creates one file of "list" input, `list.txt`, which contains three vectors (`med`, `beh` and `com`) and another file of BUGS-style "structure" input, `struct.txt`, which contains two matrices (`pscale` and `tscale`). Each of these SAS macros requires only two parameters: `FILE=` and `VAR=`. `DATA=` is optional and it defaults to the last SAS dataset created. Other optional settings are `LS=` for line length and `FORMAT=` for variable length which you can change to make your input files more readable (e.g. for dummy variables, `LS=78`, `FORMAT=1.`).

After generating posterior samples with BUGS, we want to create SAS datasets from our CODA files. Here's a code snippet:

```
goptions device=jpeg gaccess=gsasfile
          gsfmode=replace;
%coda2sas(out=post1, infile=mta.ind,
          chain=mta1.out, stats=1);
goptions device=pslepf;
%coda2sas(out=post2, infile=mta.ind,
          chain=mta2.out, stats=1, type=eps);
%_decoda(out=post3, infile=mta.ind,
          chains=2);
%bayesintervals(vars=c_1-c_3, tail=U);
```

The `CODA2SAS` macro was originally written with BUGS in mind. It can process either BUGS or WinBUGS CODA files, but doesn't handle multiple chains automatically. You have to import each of the chains manually. Set `INFILE=` to the name of your index file and `CHAIN=` to the name of your chain file. So, I imported the first chain into the SAS

dataset `post1`. This SAS dataset contains three variables `c_1-c_3` which correspond to the monitored array `c[]` where `c[1]` was translated as `c_1`, etc. But first, I set my graphics device with a `GOPTIONS` statement because when you request a summary (`STATS=1`), statistics and kernel density plots with histograms are generated. Without a `GOPTIONS` statement like the one above, the plots will be displayed on your graphics device. However, if you choose a device that is a graphics file type (like `JPEG`), then graphics files are generated for each variable (`c_1.jpg`, etc.). If a different graphics file type is declared, then you must also set the `TYPE=` option which defaults to `jpg`. For example, the second chain is imported into `post2` and *encapsulated PostScript* graphics files are created (`c_1.eps`, etc.).

Next, I read in both chains with the `_DECODA` macro. Name your index file either `NAME.ind`, `NAMEIndex.txt` or `NAME.ind.txt` and each of your chains `NAME#.out`, `NAME#.txt` or `NAME#.out.txt` respectively, where `#` is the number of your chains (1-2 in this example). If you follow this naming convention, then an unlimited number of chains are supported by specifying only `INFILE=` for the index file and `CHAINS=` for the number of chains. If the chain files follow some other naming convention, then you can specify each of them manually as `CHAIN1=` up to `CHAIN10=`. The `INFILE=` and `OUT=` parameters are required as well as either `CHAINS=` or `CHAIN1=`, etc. The posterior samples from both chains are contained in the SAS dataset `post3`. The same `CODA2SAS` comments apply to `_DECODA` with respect to variable names, `STATS=` and `TYPE=`. One additional fea-

ture is that tests and tables for location are also performed and the default location of 0 can be changed with `MU0=`. Lastly, I used the SAS macro `BAYESINTERVALS` by RD Wolfinger which is available at <http://ftp.sas.com/samples/A56648>. It constructs simultaneous intervals of the posterior for `c_1-c_3`.

There are two other SAS macros that you might find useful: `SAS2CODA` and `_LIMPORT`. `SAS2CODA` with similar syntax to `CODA2SAS` (except there is no `STATS=` option) reads a SAS dataset and creates CODA files. `_LIMPORT` creates a SAS dataset from a "list" file with two required options, `OUT=` for the new SAS dataset and `INFILE=` for the "list" file. However, importing can be tricky so a temporary SAS/IML program (which you can name with `FILE=`) is created and run automatically. If the importing fails, then you should be able to make corrections to the program and run it manually.

`CODA2SAS`, `SAS2CODA` and `BAYESINTERVALS` are self-contained and do not require any other SAS macros. However, `_DECODA`, `_LEXPOR`, `_SEXPOR` and `_LIMPORT` require other SAS macros which are also included in `decoda.zip` (currently there are 18 others, but knowledge of them is not necessary). Here are the instructions to install all SAS macros contained in `decoda.zip` (similar instructions would apply to other SAS macros):

1. `unzip decoda.zip` in a directory I'll generically call `SASMACRO` (you can call it anything you want)
2. add the following line to your SAS configuration file (`sasv7.cfg`, `sasv8.cfg` or `sasv9.cfg`)
`-insert sasautos SASMACRO`

NEWS FROM THE WORLD

by Gabriel Huerta
ghuerta@stat.unm.edu

* denotes an ISBA activity

► Events

* **Second Announcement: Second International IMS/ISBA Joint Meeting "MCMSki": The Past, Present, and Future of Gibbs Sampling.** *January 12-14 2005. Bormio, Italy (Italian Alps)*

This is second announcement of the second joint international meeting of IMS (Institute of Mathematical Statistics) and ISBA (International Society for Bayesian Analysis). Readers are encouraged to consult the conference website for full details of the main points described below:

Plenary Speakers (all confirmed)

Prof. Persi Diaconis, Stanford University
Prof. Alan Gelfand, Duke University
Prof. Sylvia Richardson, Imperial College London

Invited Sessions and Speakers (all confirmed)

Molecular Biology (organizer: Giovanni Parmigiani)
Elizabeth Garrett-Mayer, Johns Hopkins University
Marco Ramoni, Harvard Medical School
Paola Sebastiani, Boston University
Michael Ochs, Fox Chase Cancer Center

Spatial and Spatiotemporal Methods (organizer: Montserrat Fuentes)
Peter Diggle, Lancaster University
Louise Ryan, Harvard School of Public Health
Jim Zidek, University of British Columbia
Francesca Dominici, Johns Hopkins University (discussant)

MCMC in Nonparametrics (organizer: Sonia Petrone)
Gareth Roberts, Lancaster University
Peter Muller, MD Anderson Cancer Center
Eugenio Regazzini, University of Pavia
Judith Rousseau, Universite Rene Descartes (discussant)

Bioinformatics and Genetics (organizer: Steve Brooks)
Peter Donnelly, University of Oxford
Jun Liu, Harvard University
Wally Gilks, MRC Biostatistics Unit, University of Cambridge

Statistical Data Mining (organizer: Paolo Giudici)
David Madigan, Rutgers University
Andrew Nobel, University of North Carolina
Paolo Giudici, University of Pavia
Greg Ridgeway, RAND Corporation (discussant)

MCMC Algorithms and Software (organizers: Anto Mira and Brad Carlin)
Luke Tierney, University of Iowa
Xiao-Li Meng, Harvard University
David Spiegelhalter, MRC Biostatistics Unit, University of Cambridge
Nicky Best, Imperial College London
Kerrie Mengersen, University of Newcastle (discussant)

Contributed poster sessions

All participants are encouraged to contribute to one of the two poster sessions, to be held in the Rezia Hotel. Instructions and a template for abstract submission may be found on the conference website.

Registration, accommodations and bus transport

Again, this information is now all on the website. Registration fees include all conference materials, refreshments during breaks, and the closing conference banquet (a multicourse meal with wine and dessert). Payment may be made by credit card via our secure web server operated by the IMS. You will be asked what day and time you plan to arrive; we realize it's early to ask this, but if you can give us some idea it will help us with respect to planning the bus schedule. Hotel reservations must be made separately by direct email contact.

Junior investigator travel support

While we cannot yet say for certain, we anticipate limited financial support for junior investigators presenting posters from the NSF and perhaps other granting agencies. NSF requires such investigators to be from US institutions (though not necessarily US citizens) and within 5 years of the PhD. Interested persons should indicate this when they register for the meeting and submit their abstracts. A recent CV and a *brief* note from your academic

supervisor is also required.
Booksellers and other vendors

Will be located outside the main lecture hall in the conference center; persons interested in having a kiosk should contact one of the co-organizers for details.

Recreational opportunities

The conference website has information on both the ski and spa offerings, which will be available at reduced rates during the afternoon breaks to conference participants and their accompanying persons. Details on the "Tweedie Cup" ski race will also be posted as they develop.

Post-Banquet Cabaret

While post-banquet cabarets have a long and celebrated history in international Bayesian conference lore, IMS-ISBA No.2 offers the opportunity to establish this tradition in a much more ecumenical setting. All persons having *any* relevant talent whatsoever (singing, playing, acting, juggling, magic tricks, jeketelling, etc.) are strongly encouraged to contact one of the co-organizers. No audition is necessary.

Breaking news: Pre-conference "satellite" meeting: "Advances in Adaptive MCMC Algorithms", Bormio (Italy), January 10-11, 2005

Organizers: Ch. Andrieu (Dept. Math and Stat., U. Bristol, UK)
H. Haario (Dept. Math. U. Helsinki, Finland)
C.P. Robert (Dept. Applied Math., U. Paris Dauphine, France)

The focus of this satellite meeting is to discuss the theoretical tools necessary for the development of adaptive MCMC algorithms and the barriers to the dissemination of such algorithms in more realistic settings, as well as set up a program for future research. Talks and open forums will center on four topics:

(1) Ergodicity of non-homogeneous Markov chains (2) Tools and advances in stochastic optimization (3) Non Markovian alternatives (4) Implementation specificities of adaptive algorithms

For more information please contact Prof. Robert: christian_robert@ensae.fr

Please direct all further IMS/ISBA No. 2 questions to one of the two co-chairs:

Brad Carlin, University of Minnesota: brad@biostat.umn.edu
Antonietta Mira, University of Insubria: antonietta.mira@uninsubria.it
Site: <http://alien.eco.uninsubria.it/IMS-ISBA-05/>

*** COBAL II Announcement February 6-10, 2005.**
San Jose del Cabo, Mexico.

Following up on the success of COBAL I (Ubatuba, Brazil; 2002), the Second Latin American Congress on Bayesian Statistics (COBAL II) will be held February 6-10, 2005 at Hotel Presidente Intercontinental, Los Cabos in San Jose del Cabo, Baja California, Mexico. This meeting will consist of invited talks and poster sessions. E-mail: cobal2@sigma.iimas.unam.mx

San Jose del Cabo is a small town located at the southern end of the Baja California peninsula. San Jose is an authentic Mexican town, settled near the ocean along the San Jose Estuary, and on nearby hills. The logs of Spanish galleons in the 1500's refer to the estuary as "aguada segura", or "safe watering place". San Jose was an important stop for sailors low on provisions. Pirates landed here for fresh water, and hid nearby to ambush Spanish galleons, loaded with Mexican gold and silver, sailing from Acapulco to the Phillippines to buy silks and spices.

Today, San Jose is a busy center for commerce, agriculture, government, fishing and tourism. The influx of foreign (and Mexican) tourists, retirees and investors in recent years has resulted in an economic boom and rapid growth that continues today. As an additional and unique attraction, each year, in February, one of nature's most delightful signs of seasonal change can be witnessed – the migration of hundreds of whales from the frigid waters of the arctic to the warm, calm waters surrounding the Baja Peninsula.

San Jose del Cabo airport is an international terminal and you can flight directly from a number of US cities, as for example Houston, Denver and Atlanta. There are also connections from Mexico City and many other Mexican airports.

Our meeting has been organized as a residential event and its venue will be the Presidente Intercontinental Hotel . Accommodation costs are included (4 nights) in the conference fee.

We strongly encourage you to visit the web site of the congress www.dpye.iimas.unam.mx/cobal2.htm where you will find not only more details on the scientific organization of event but also useful information for your travel to San Jose del Cabo.

Further information:

Web Site: www.dpye.iimas.unam.mx/cobal2.htm
E-mail: cobal2@sigma.iimas.unam.mx

Regards, Organizing Committee COBAL 2.

*** Fourth Workshop on BAYESIAN INFERENCE IN STOCHASTIC PROCESSES.** June 2-4 2005. Villa Monastero, Varenna (LC), Italy

Call for Interest

The workshop follows the ones held in Madrid in 1998, in Varenna in 2001 and La Manga in 2003 (see <http://www.upct.es/bisp03/> for the last one).

In this workshop, we will bring together experts in the field to review, discuss and explore directions of development of Bayesian Inference in Stochastic Processes and in the use of Stochastic Processes for Bayesian Inference. There will be sessions on Markov processes, state-space models, spatial, empirical, birth-death and branching processes. Theoretical and applied contributions (for example queueing, population modelling, signal processing) are both welcome.

The workshop will thus be of interest to those working in both Bayesian Inference and Stochastic Processes.

The workshop will be held in an informal environment (a '700 villa on the lake of Como) to encourage discussion and promote further research in these fields.

Organisation: CNR IMATI, Milano Dept. (<http://www.mi.imati.cnr.it>)

For inquiries please contact bisp4@mi.imati.cnr.it

Site: <http://www.mi.imati.cnr.it/conferences/bisp4.html>



INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

Executive Committee

President: Jim Berger
Past President: Ed George
President Elect: Sylvia Richardson
Treasurer: Peter Müller
Executive Secretary:
Deborah Ashby

Program Council

Chair: Alan Gelfand
Vice Chair: José Bernardo
Past Chair: Luis Pericchi

Web page: www.bayesian.org

Board Members

Brad Carlin, Merlise Clyde, Michael Goldstein, David Higdon, Pilar Iglesias, Jun Liu, David Madigan, Sonia Petrone, Christian Robert, Fabrizio Ruggeri, Marina Vannucci, Robert Wolpert

EDITORIAL BOARD

Editor

Hedibert Lopes <hedibert.lopes@gsb.uchicago.edu>

Associate Editors

Annotated Bibliography
Duncan Fong <i2v@psu.edu>

Applications
Bruno Sansó <bruno@ams.ucsc.edu>

Interviews
Brunero Liseo <brunero.liseo@uniroma1.it>

News from the World
Gabriel Huerta <ghuerta@stat.unm.edu>

Software Review
Viridiana Lourdes <vlourdes@itam.mx>

Student's Corner
Lilla Di Scala <lilla@dimat.unipv.it>

Bayesian History
Sérgio Weschler <sw@ime.usp.br>
Antonio Pievatolo <marco@mi.imati.cnr.it>