Bayesian Analysis Users Guide
Release 4.00, Manual Version 1

G. Larry Bretthorst
Biomedical MR Laboratory
Washington University School Of Medicine,
Campus Box 8227
Room 2313, East Bldg.,
4525 Scott Ave.
St. Louis MO 63110
http://bayes.wustl.edu
Email: larry@bayes.wustl.edu

October 21, 2016

Contents

List of Figures

List of Tables

Appendix A

Ascii Data File Formats

Ascii data files are used through out the entire Bayesian Analysis software. Often they are used for simple things like input to various packages. Sometimes they are used to loading Abscissa into plotting routines. And sometimes they are used to generate Fid and Image files. In all cases the exact file formats vary depending on the type of data and the package that is loading the Ascii file. This Appendix is a description of these file formats and how they are used. Most of the time the file formats are pretty simple and rather self explanatory and we will list a few of these shortly. However, in a few cases the actual file format of the Ascii data can be complicated. We give these more complicated file formats in this Appendix. For now here is a list of some of the more simple Ascii file formats:

Images can be loaded in Ascii. The interface Files/Load Image/Single-Column text file widget expects the data to be single column Ascii with each row in the image stacked one line after another.

FID data can be loaded as Ascii data. In this case one complex (real and imaginary) number are expected on each like of the FID. This type of data is used in the Files/Load Spectroscopic Fid/Text File widget.

Ascii Input Files for most packages, like an exponential package, allow Ascii data to be loaded. In these simple cases, the files are simply two column Ascii: one abscissa and one data column. However, see below because there are major exceptions to this rule.

A.1 Ascii Input Data Files

The format of a general input Ascii file used in the Bayesian Analysis software is shown in Fig. A.1. Each line in an input Ascii file consist of three parts. This is illustrated by the double vertical lines in the figure. The first part is a single column plot abscissa, and as its name implies, this is the abscissa used in plotting the Ascii data. All plots of the Ascii data done by the interface are of the data verses the plot abscissa. If there are multiple data columns, then multiple plots are drawn by the interface. However, all of these data plots use the input plot abscissa on the current data set.

The second part of an Ascii input file, is the data and as indicated in Fig. A.1, there can be up to N data columns, where N is specified by the model, i.e., it is user defined. The number of these

Figure A.1: Ascii Data File Format

Plot Abscissa	Data Col 1	...	Data Col N	Abscissa Col 1	$\cdots$	Abscissa Col M
1	25.1	$\cdots$	30.2	1	$\cdots$	3
2	15.1	$\cdots$	10.2	2	$\cdots$	-1
$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Figure A.1: This table shows a schematic of a general input Ascii data file. The first column is simply a plot variable and in the case of a single column abscissa, this variable should be the abscissa. The columns labeled "Data Col 1" though "Data Col N" are the N input data columns. The number of these column is implicit in the package model or it is defined explicitly in the model parameter file. Finally, when multiple abscissa's are present, these abscissa follow the data columns. Here these M columns are labeled "Abscissa Col 1" through "Abscissa Col M". If there is only a single abscissa, then these abscissa columns are not present.

columns is often implicit in the nature of the package. For example, an inversion recovery data set would require only a single data column However, a magnetization transfer packages might require real and imaginary data so the data section might have two columns. User defined models can be written that use any number of input data columns. In Fig. A.1 these data columns are labeled as "Data Col 1" through "Data Col N" where there are N data columns.

Finally the computational abscissa section is present only when the package expects multicolumn abscissa. For example, in a diffusion tensor analysis each abscissa consist of a 3 dimensional B vector, so 3 abscissa are needed. When a model runs that uses a computational abscissa, the abscissa passed to the model program contains as many columns as specified by the model. For example, a diffusion tensor analysis using a B vector would be a 3 by n abscissa, where 3 is the number of abscissa columns and n is the number of data values in the current data set.

In Ascii packages multiple data sets can be loaded and analyzed jointly. Each data set can have a differing number of data values and different abscissa values. However, the number of abscissa columns is fixed for a given analysis. When the posterior probability is computed the model program must generate the model given the current parameters and abscissa. After this information is generated, the posterior probability for the current data set is computed and used in the Bayesian calculations. So while you could have different abscissa values in the different data sets, all data sets must use the same number of abscissa columns.

A.2 Ascii Image File Formats

The previous section describes the file format for Ascii data files that are used as input to the various packages. However, Ascii images can also be loaded and in this case there are three distinct image formats, one single column, one multicolumn and one k-space Ascii format.

In the case of single column Ascii images, the images are stacked in the Ascii file one pixel at a time. If the image consists of N rows by M columns then the interface expects rows 1 (all M values) followed by row 2, etc until all N rows are read. Images can be stacked by slice or element number and you can specify in the popup how they are stacked so the outer loops can be either slice number or array element number.

Multicolumn Ascii files can also be loaded. In this case all elements in a give row are read from record one of the input file. Record 2 corresponds to the second row in the image, etc. The ordering of the images by slice and element is again under user control and can be specified when the images are loaded.

Finally, images can be complex k-spaced data. In this case, the complex k-space data is two column Ascii data. Each row of the complex data corresponds to one readout and readouts are stacked one after another. This complex k-space data is converted into a Varian fid. This fid is then Fourier transformed, phased, and displayed as a complex image. If multiple k-space images are present, the slice and element orders are again under user control. For k-space fid data the output from the interface is the real, imaginary and absolute value images.

All images are whether or not they are input as Ascii files, k-space files, or any other formats are copied into your current working directory. These images are located in BayesHome/WorkDir/images, where "BayesHome" is your current Bayes home directory, "WorkDir" is your current working directory, and "images" is the images subdirectory. Loaded images preserve the name of the name of the input image and duplicates will overwrite the image directory. However, loaded k-space data are first converted into a Varian fid. This fid is located in the BayesHome/WorkDir/Image.fid subdirectory. After this conversion, the k-space data is then Fourier transformed and displayed as an image. These images are output with the name "LoadedImage_Real.4dfp.ima," "LoadedImage_Imag.fdfp.ima," and "LoadedImage_Abs.4dfp.ima" for the real, imaginary and absolute value images respectively.

In all of these cases, an abscissa file will almost certainly have to be loaded and in the case of k-space data is required for the conversion. The next section briefly discusses the format of an abscissa file used with images.

A.3 The Abscissa File Format

If an image is arrayed, the values of the array variable must be specified in an abscissa file before any image processing can be done on the image. Additionally, when k-space images are loaded, the abscissa must be available to create the Varian fid file. These abscissa values are given in multicolumn Ascii files, one column for each abscissa column. Unusually abscissa are one dimensional like, for example, in an inversion recovery experiment. In which case the abscissa file is a simple single column Ascii file. But sometimes they can be much more complicated, for example, a B matrix diffusion experiment would have to have a 6 column abscissa, one entry for each of the 6 independent B values in a diffusion experiment. In all cases the actual meaning of the columns in an abscissa are specified by the model used in processing the image data. For example a diffusion tensor image would have an abscissa consisting of the gradient or B values used in the experiment. These gradient or B values could be one, two, etc. up to 6 dimensional depending on the problem. So the abscissa file could be one column, as in diffusion along a single direction, two column, for diffusion in a plane, all the way up to a 6 column abscissa for a full B matrix diffusion tensor experiment.

Bibliography

[1] Rev. Thomas Bayes (1763), "An Essay Toward Solving a Problem in the Doctrine of Chances," *Philos. Trans. R. Soc. London*, **53,** pp. 370-418; reprinted in *Biometrika*, **45**, pp. 293-315 (1958), and *Facsimiles of Two Papers by Bayes,* with commentary by W. Edwards Deming, New York, Hafner, 1963.

[2] G. Larry Bretthorst (1988), "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics,* **48**, J. Berger, S. Fienberg, J. Gani, K. Krickenberg, and B. Singer (eds), Springer-Verlag, New York, New York.

[3] G. Larry Bretthorst (1990), "An Introduction to Parameter Estimation Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods,* Dartmouth College 1989, P. Fougère *ed.,* pp. 53-79, Kluwer Academic Publishers, Dordrecht the Netherlands.

[4] G. Larry Bretthorst (1990), "Bayesian Analysis I. Parameter Estimation Using Quadrature NMR Models" *J. Magn. Reson.*, **88**, pp. 533-551.

[5] G. Larry Bretthorst (1990), "Bayesian Analysis II. Signal Detection And Model Selection" *J. Magn. Reson.*, **88**, pp. 552-570.

[6] G. Larry Bretthorst (1990), "Bayesian Analysis III. Examples Relevant to NMR" *J. Magn. Reson.*, **88**, pp. 571-595.

[7] G. Larry Bretthorst (1991), "Bayesian Analysis. IV. Noise and Computing Time Considerations," *J. Magn. Reson.*, **93**, pp. 369-394.

[8] G. Larry Bretthorst (1992), "Bayesian Analysis. V. Amplitude Estimation for Multiple Well-Separated Sinusoids," *J. Magn. Reson.*, **98**, pp. 501-523.

[9] G. Larry Bretthorst (1992), "Estimating The Ratio Of Two Amplitudes In Nuclear Magnetic Resonance Data," in *Maximum Entropy and Bayesian Methods,* C. R. Smith *et al.* (eds.), pp. 67-77, Kluwer Academic Publishers, the Netherlands.

[10] G. Larry Bretthorst (1993), "On The Difference In Means," in Physics & Probability Essays in honor of Edwin T. Jaynes, W. T. Grandy and P. W. Milonni (eds.), pp. 177-194, Cambridge University Press, England.

[11] G. Larry Bretthorst (1996), "An Introduction To Model Selection Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, ed., pp. 1-42, Kluwer Academic Publishers, Printed in the Netherlands.

[12] G. Larry Bretthorst (1999), "The Near-Irrelevance of Sampling Frequency Distributions," in *Maximum Entropy and Bayesian Methods,* W. von der Linden *et al.* (eds.), pp. 21-46, Kluwer Academic Publishers, the Netherlands.

[13] G. Larry Bretthorst (2001), "Nonuniform Sampling: Bandwidth and Aliasing," in *Maximum Entropy and Bayesian Methods in Science and Engineering,* Joshua Rychert, Gary Erickson and C. Ray Smith *eds.*, pp. 1-28, American Institute of Physics, USA.

[14] G. Larry Bretthorst, Christopher D. Kroenke, and Jeffrey J. Neil (2004), "Characterizing Water Diffusion In Fixed Baboon Brain," in Bayesian Inference And Maximum Entropy Methods In Science And Engineering, Rainer Fischer, Roland Preuss and Udo von Toussaint *eds.*, AIP conference Proceedings, **735**, pp. 3-15.

[15] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J.H. Ackerman (2005), "Exponential parameter estimation (in NMR) using Bayesian probability theory," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 55-63.

[16] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J. H. Ackerman (2005), "Exponential model selection (in NMR) using Bayesian probability theory," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 64-72.

[17] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J.H. Ackerman (2005), "How accurately can parameters from exponential models be estimated? A Bayesian view," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 73-83.

[18] G. Larry Bretthorst, W. C. Hutton, J. R. Garbow, and Joseph J. H. Ackerman (2008), "High Dynamic Range MRS Time-Domain Signal Analysis," *Magn. Reson. in Med.*, **62**, pp. 1026-1035.

[19] V. Chandramouli, K. Ekberg, W. C. Schumann, S. C. Kalhan, J. Wahren, and B. R. Landau (1997), "Quantifying gluconeogenesis during fasting," *American Journal of Physiology*, **273**, pp. H1209-H1215.

[20] R. T. Cox (1961), "The Algebra of Probable Inference," Johns Hopkins Univ. Press, Baltimore.

[21] André d'Avignon, G. Larry Bretthorst, Marlyn Emerson Holtzer, and Alfred Holtzer (1998), "Site-Specific Thermodynamics and Kinetics of a Coiled-Coil Transition by Spin Inversion Transfer NMR," *Biophysical Journal*, **74**, pp. 3190-3197.

[22] André d'Avignon, G. Larry Bretthorst, Marlyn Emerson Holtzer, and Alfred Holtzer (1999), "Thermodynamics and Kinetics of a Folded-Folded Transition at Valine-9 of a GCN4-Like Leucine Zipper," *Biophysical Journal*, **76**, pp. 2752-2759.

[23] David Freedman, and Persi Diaconis (1981), "On the histogram as a density estimator: L_2 theory," *Zeitschrift f¨r Wahrscheinlichkeitstheorie und verwandte Gebiete*, **57**, 4, pp. 453-476.

[24] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (1996), "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.

[25] Paul M. Goggans, and Ying Chi (2004), "Using Thermodynamic Integration to Calculate the Posterior Probability in Bayesian Model Selection Problems," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 23rd International Workshop*, **707**, pp. 59-66.

[26] Marlyn Emerson Holtzer, G. Larry Bretthorst, D. André d'Avignon, Ruth Hogue Angelette, Lisa Mints, and Alfred Holtzer (2001), "Temperature Dependence of the Folding and Unfolding Kinetics of the GCN4 Leucine Lipper via 13C alpha-NMR," *Biophysical Journal*, **80**, pp. 939-951.

[27] E. T. Jaynes (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, pp. 227-241; reprinted in [30].

[28] E. T. Jaynes (1978), "Where Do We Stand On Maximum Entropy?" in *The Maximum Entropy Formalism*, R. D. Levine and M. Tribus *Eds.*, pp. 15-118, Cambridge: MIT Press, Reprinted in [30].

[29] E. T. Jaynes (1980), "Marginalization and Prior Probabilities," in *Bayesian Analysis in Econometrics and Statistics*, A. Zellner *ed.,* North-Holland Publishing Company, Amsterdam; reprinted in [30].

[30] E. T. Jaynes (1983), "Papers on Probability, Statistics and Statistical Physics," a reprint collection, D. Reidel, Dordrecht the Netherlands; second edition Kluwer Academic Publishers, Dordrecht the Netherlands, 1989.

[31] E. T. Jaynes (1957), "How Does the Brain do Plausible Reasoning?" unpublished Stanford University Microwave Laboratory Report No. 421; reprinted in *Maximum-Entropy and Bayesian Methods in Science and Engineering* **1,** pp. 1-24, G. J. Erickson and C. R. Smith *Eds.,* 1988.

[32] E. T. Jaynes (2003), "Probability Theory—The Logic of Science," edited by G. Larry Bretthorst, Cambridge University Press, Cambridge UK.

[33] Sir Harold Jeffreys (1939), "Theory of Probability," Oxford Univ. Press, London; Later editions, 1948, 1961.

[34] John G. Jones, Michael A. Solomon, Suzanne M. Cole, A. Dean Sherry, and Craig R. Malloy (2001) "An integrated ^{2}H and ^{13}C NMR study of gluconeogenesis and TCA cycle flux in humans," *American Journal of Physiology, Endocrinology, and Metabolism,* **281,** pp. H848-H856.

[35] John Kotyk, N. G. Hoffman, W. C. Hutton, G. Larry Bretthorst, and J. J. H. Ackerman (1992), "Comparison of Fourier and Bayesian Analysis of NMR Signals. I. Well-Separated Resonances (The Single-Frequency Case)," *J. Magn. Reson.*, **98**, pp. 483–500.

[36] Pierre Simon Laplace (1814), "A Philosophical Essay on Probabilities," John Wiley & Sons, London, Chapman & Hall, Limited 1902. Translated from the 6th edition by F. W. Truscott and F. L. Emory.

[37] N. Lartillot, and H. Philippe (2006), "Computing Bayes Factors Using Thermodynamic Integration," *Systematic Biology*, **55** (2), pp. 195-207.

[38] D. Le Bihan, and E. Breton (1985), "Imagerie de diffusion in-vivo par rsonance," Comptes rendus de l'Acadmie des Sciences (Paris), **301** (15), pp. 1109-1112.

[39] N. R. Lomb (1976), "Least-Squares Frequency Analysis of Unevenly Spaced Data," *Astrophysical and Space Science*, **39**, pp. 447-462.

[40] T. J. Loredo (1990), "From Laplace To SN 1987A: Bayesian Inference In Astrophysics," in *Maximum Entropy and Bayesian Methods,* P. F. Fougere (ed), Kluwer Academic Publishers, Dordrecht, The Netherlands.

[41] Craig R. Malloy, A. Dean Sherry, and Mark Jeffrey (1988), "Evaluation of Carbon Flux and Substrate Selection through Alternate Pathways Involving the Citric Acid Cycle of the Heart by 13C NMR Spectroscopy," *Journal of Biological Chemistry*, **263** (15), pp. 6964-6971.

[42] Craig R. Malloy, Dean Sherry, and Mark Jeffrey (1990), "Analysis of tricarboxylic acid cycle of the heart using ^{13}C isotope isomers," *American Journal of Physiology*, **259**, pp. H987-H995.

[43] Lawrence R. Mead and Nikos Papanicolaou, "Maximum entropy in the problem of moments," *J. Math. Phys.* **25**, 2404–2417 (1984).

[44] K. Merboldt, Wolfgang Hanicke, and Jens Frahm (1969), "Self-diffusion NMR imaging using stimulated echoes," Journal of Magnetic Resonance, **64** (3), pp. 479-486.

[45] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953), "Equation of State Calculations by Fast Computing Machines," Journal of Chemical Physics. The previous link is to the Americain Institute of Physics and if you do not have access to Science Sitations you many not be able to retrieve this paper.

[46] Radford M. Neal (1993), "Probabilistic Inference Using Markov Chain Monte Carlo Methods," technical report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.

[47] Jeffrey J. Neil, and G. Larry Bretthorst (1993), "On the Use of Bayesian Probability Theory for Analysis of Exponential Decay Data: An Example Taken from Intravoxel Incoherent Motion Experiments," *Magn. Reson. in Med.*, **29**, pp. 642–647.

[48] H. Nyquist (1924), "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, **3**, pp. 324-346.

[49] H. Nyquist (1928), "Certain Topics in Telegraph Transmission Theory," *Transactions AIEE*, **3**, pp. 617-644.

[50] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery (1992), "Numerical Recipes The Art of Scientific Computing Second Edition," Cambridge University Press, Cambridge UK.

[51] Emanuel Parzen (1962), "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics* **33**, 1065–1076

[52] Karl Pearson (1895), "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," *Phil. Trans. R. Soc. A* **186**, 343–326.

[53] Murray Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics* **27**, 832–837 (1956).

[54] Jeffery D. Scargle (1981), "Studies in Astronomical Time Series Analysis I. Random Process In The Time Domain," *Astrophysical Journal Supplement Series*, **45**, pp. 1-71.

[55] Jeffery D. Scargle (1982), "Studies in Astronomical Time Series Analysis II. Statistical Aspects of Spectral Analysis of Unevenly Sampled Data," *Astrophysical Journal*, **263**, pp. 835-853.

[56] Jeffery D. Scargle (1989), "Studies in Astronomical Time Series Analysis. III. Fourier Transforms, Autocorrelation Functions, and Cross-correlation Functions of Unevenly Spaced Data," *Astrophysical Journal*, **343**, pp. 874-887.

[57] Arthur Schuster (1905), "The Periodogram and its Optical Analogy," *Proceedings of the Royal Society of London*, **77**, p. 136-140.

[58] Claude E. Shannon (1948), "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, **27**, pp. 379-423.

[59] John E. Shore, and Rodney W. Johnson (1981), "Properties of cross-entropy minimization," *IEEE Trans. on Information Theory*, **IT-27,** No. 4, pp. 472-482.

[60] John E. Shore and Rodney W. Johnson (1980), "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. on Information Theory*, **IT-26** (1), pp. 26-37.

[61] Devinderjit Sivia, and John Skilling (2006), "Data Analysis: A Bayesian Tutorial," Oxford University Press, USA.

[62] Edward O. Stejskal and Tanner, J. E. (1965), "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient." Journal of Chemical Physics, **42** (1), pp. 288-292.

[63] D. G. Taylor and Bushell, M. C. (1985), "The spatial mapping of translational diffusion coefficients by the NMR imaging technique," Physics in Medicine and Biology, **30** (4), pp. 345-349.

[64] Myron Tribus (1969), "Rational Descriptions, Decisions and Designs," Pergamon Press, Oxford.

[65] P. M. Woodward (1953), "Probability and Information Theory, with Applications to Radar," McGraw-Hill, N. Y. Second edition (1987); R. E. Krieger Pub. Co., Malabar, Florida.

[66] Arnold Zellner (1971), "An Introduction to Bayesian Inference in Econometrics," John Wiley and Sons, New York.