Introduction to Bayesian Inference: Selected Resources

Tom Loredo
Cornell Center for Astrophysics and Planetary Science
http://www.astro.cornell.edu/staff/loredo/bayes/

CASt Summer School — 7–8 June 2017

Books by physicists and astronomers

- *Probability Theory: The Logic of Science* (PTLOS)
  Edwin T. Jaynes; ed. G. Larry Bretthorst [http://bayes.wustl.edu/]
  [Cambridge U. Press]
  Jaynes worked on this book for over 30 years; it was unfinished at his death in 1998, but Bretthorst thankfully assembled the book from his last draft chapters. Provides the best (and lengthiest) coverage of foundations and fundamentals for a physical scientist audience. It dates from before the development of modern computational tools, and is thus not the most practical text.
  See reviews by: Persi Diaconis (theoretical & applied statistics), Anton Garrett (physics), Terry Fine (applied math, philosophy), Will Faris (for AMS).
  Diaconis: "There are many places in which I want to yell at him. He's so full of himself. That's what makes the book so terrific. It's the real thing—the best introduction to Bayesian statistics that I know. Go take a look for yourself."
- *Bayesian Logical Data Analysis for the Physical Sciences, A Comparative Approach with Mathematica Support*
  Phil Gregory [Cambridge U. Press (2010)]
  Could be regarded as a practical companion to PTLOS; adopts similar point of view but focuses on applications, including basic coverage of MCMC. Some comparison with frequentist approaches.
- *Data Analysis: A Bayesian Tutorial*
  Devinder Sivia, John Skilling [Oxford U. Press (2006)]
  The most accessible book on Bayesian methods by physical scientists; somewhat idiosyncratic coverage of computational methods.

- *Bayesian Probability Theory: Applications in the Physical Sciences*
  Wolfgang von der Linden, Volker Dose, Udo von Toussaint
  [Cambridge U. Press, coming July 2014]
  Authors are highly-regarded pioneers of application of Bayesian methods to problems in plasma physics and other areas. Some weaknesses on theory/fundamental topics, but numerous very good examples from physics.
- *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*
  Zeljko Ivezić, Andrew Connolly, Jacob VanderPlas, Alexander Gray
  [Princeton U. Press]
  Balanced coverage of frequentist and Bayesian methods, mostly in the context of analyzing large survey datasets. Extensive accompanying Python software, datasets, and reproducible analyses.
- *Bayesian Methods for the Physical Sciences*
  Stefano Andreon, Brian Weaver [Springer; authors' site]
  New (2015) book by astronomers, highlighting use of the JAGS probabilistic programming language. See the somewhat mixed review by astronomer David Hogg.
- *Bayesian Models for Astrophysical Data Using R, JAGS, Python, and Stan*
  By statistician Joseph Hilbe and astronomers Rafael de Souza and Emille E. O. Ishida [Cambridge U. Press]

- *Information Theory, Inference, and Learning Algorithms*
  David MacKay [Cambridge U. Press, 2003; free PDF/DJVU at MacKay's site]
  By a physicist-turned-statistician/information theorist. An extremely original and influential account of ideas underlying statistics, machine learning, signal processing, and communication, from a Bayesian viewpoint. A strong emphasis on information theory and coding problems makes it not the most straightforward introduction for a data analyst, yet it has exceptionally clear coverage of model comparison, information-based experimental design, neural networks, and Monte Carlo methods (including MCMC).
- *Bayesian Methods in Cosmology*
  Ed. by Michael Hobson et al. [Cambridge U. Press (2010)]
  Chapters by multiple authors and thus with varying quality and notation.

Tutorials aimed at physical scientists

See links collected at the Bayesian inference for the physical sciences (BIPS) web site. Note that this site is not regularly updated; some noteworthy recent articles include:

- "Bayesian Methods in Cosmology" by Roberto Trotta — ADS, arXiv:1701.01467
- "Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy" by Sanjib Sharma — arXiv:1706.01629

Selected Bayesian statistics books

- *Doing Bayesian Data Analysis*
  John K. Kruschke [author's book site]
  Known as "the dog book," for the illustration of dogs on the cover, it offers an exceptionally clear, thorough, and accessible introduction to Bayesian concepts and computational techniques. I recommend this to beginning students. Be sure to get the 2nd edn., which switches from BUGS to JAGS and Stan as computational tools.
- *Bayesian Data Analysis* (BDA)
  Andrew Gelman et al. [CRC Press (3rd edn. 2013)]
  Probably the most influential and widely-used Bayesian text by statisticians. Both broad and deep, including coverage of multilevel modeling, nonparametric Bayes, model testing, and modern computational methods.
- *Handbook of Markov Chain Monte Carlo*
  Ed. by Brooks, Gelman, Jones, Meng [CRC Press (2011)]
  Accessible, authoritative coverage of a wide range of MCMC techniques, including good coverage of output analysis. Selected chapters online.
- *Bayesian Methods for Data Analysis*
  Bradley Carlin & Thomas Louis [CRC Press (3rd edn. 2008)
  Earlier editions were titled, "Bayes and Empirical Bayes Methods for Data Analysis," reflecting the book's particularly strong coverage of empirical/hierarchical Bayesian modeling (multilevel modeling). See Gelman's comparison of BDA and Carlin & Louis.

There are many other excellent Bayesian texts by statisticians; this brief, idiosyncratic list just scratches the surface.

Tools for Computational Bayes

Astronomer/Physicist Tools

- **BIE** `http://www.astro.umass.edu/~weinberg/BIE/`
  Bayesian Inference Engine: General framework for Bayesian inference, tailored to astronomical and earth-science survey data. Built-in database capability to support analysis of terabyte-scale data sets. Inference is by Bayes via MCMC. Documentation limited.
- **AstroML** `http://www.astroml.org/`
  Python package supporting machine learning and statistical inference for analyzing astronomical data. Built in part to support the book, "Statistics, Data Mining, and Machine Learning in Astronomy;" it includes modules supporting Bayesian calculations from the book. Well-maintained, well-documented.
- **CosmoMC** `http://cosmologist.info/cosmomc/`
  Parameter estimation for cosmological models using CMB, etc., via MCMC
- **DNest4** `https://github.com/eggplantbren/DNest4`
  Posterior sampling and marginal likelihoods via diffusive nested sampling
- **MultiNest** `http://ccpforge.cse.rl.ac.uk/gf/project/multinest/`
  Bayesian inference via an approximate implementation of the nested sampling algorithm
- **PolyChord** `https://ccpforge.cse.rl.ac.uk/gf/project/polychord/`
  "Next generation" nested sampling

- **emcee** http://dan.iel.fm/emcee/
  Python implementation of an ensemble MCMC sampler (no diagnostics—be sure to find them elsewhere!)
- **extreme-deconvolution**
  http://code.google.com/p/extreme-deconvolution/
  Multivariate density estimation with measurement error, via a multivariate normal finite mixture model; partly Bayesian; Python & IDL wrappers
- **George** http://dan.iel.fm/george/
  Fast Gaussian process implementation, for nonparametric Bayesian regression.
- **ExoFit** http://www.homepages.ucl.ac.uk/~ucapola/exofit.html
  Adaptive MCMC for fitting exoplanet RV data
- **XSpec** http://heasarc.nasa.gov/xanadu/xspec/
  Includes some basic MCMC capability
- **CIAO/Sherpa** http://cxc.harvard.edu/sherpa/
  On/off marginal likelihood support, and Bayesian Low-Count X-ray Spectral (BLoCXS) analysis via MCMC via the **pyblocxs** extension
  https://github.com/brefsdal/pyblocxs
- **root/RooStats** https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome
  Statistical tools for particle physicists; Bayesian support being incorporated
- **CDF Bayesian Limit Software**
  http://www-cdf.fnal.gov/physics/statistics/statistics_software.html
  Limits for Poisson counting processes, with background & efficiency uncertainties

- **CUBA** `http://www.feynarts.de/cuba/`
  Multidimensional integration via adaptive cubature, adaptive importance sampling & stratification, and QMC (C/C++, Fortran, and Mathematica; R interface also via 3rd-party R2Cuba)
- **Cubature** `http://ab-initio.mit.edu/wiki/index.php/Cubature`
  Subregion-adaptive cubature in C, with a 3rd-party R interface; intended for low dimensions (< 7)
- **APEMoST** `http://apemost.sourceforge.net/doc/`
  Automated Parameter Estimation and Model Selection Toolkit in C, a general-purpose MCMC environment that includes parallel computing support via MPI; motivated by asteroseismology problems
- **SuperBayeS** `http://www.superbayes.org/`
  Bayesian exploration of supersymmetric theories in particle physics using the MultiNest algorithm; includes a MATLAB GUI for plotting
- **Inference** Forthcoming at `http://inference.astro.cornell.edu/`
  Python package targeting statistical inference problems arising in the physical sciences; several self-contained Bayesian modules; Parametric Inference Engine

Python

- **PyStan** `https://pystan.readthedocs.io/`
  Python interface to the Stan probabilistic programming language, for partly automated posterior sampling for graphical (hierarchical) models. See also TL's `StanFitter` for a more Pythonic interface.
- **PyMC** `http://code.google.com/p/pymc/`
  A framework for MCMC via Metropolis-Hastings; also implements Kalman filters and Gaussian processes. Targets biometrics, but is general. Includes output analysis tools.
- **emcee** `http://dan.iel.fm/emcee/current/`
  Python implementation of an ensemble-based, affine invariant MCMC algorithm, by astronomer Daniel Foreman-Mackey.
- **Monte Python** `http://baudren.github.io/montepython.html`
  A a Monte Carlo code for Cosmological Parameter extraction.
- **SimPy** `http://simpy.sourceforge.net/`
  SimPy (rhymes with "Blimpie") is a process-oriented public-domain package for discrete-event simulation.
- **rpy2** `http://rpy2.readthedocs.io/`
  Call R from Python; see the CRAN Bayesian Task View for Bayesian resources. Also see **RSPython** `https://web.archive.org/web/20151130002540/http://www.omegahat.org/RSPython`, with bi-directional communication between Python and R (abandoned?)
- **Inference** Forthcoming package by TL, including a Parametric Inference Engine (PIE) module implementing various Bayesian computation methods.

R packages and interfaces

- **CRAN Bayesian task view**
  `http://cran.r-project.org/web/views/Bayesian.html`
  Overview of many R packages implementing various Bayesian models and methods; pedagogical packages; packages linking R to other Bayesian software (BUGS, JAGS)
- **BOA** `http://www.public-health.uiowa.edu/boa/`
  Bayesian Output Analysis: Convergence diagnostics and statistical and graphical analysis of MCMC output; can read BUGS output files.
- **CODA** `http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html`
  Convergence Diagnosis and Output Analysis: Menu-driven R/S plugins for analyzing BUGS output
- **LearnBayes**
  `http://cran.r-project.org/web/packages/LearnBayes/index.html`
  Companion software for the introductory book, *Bayesian Computation With R* by Jim Albert
- **R2Cuba** `http://w3.jouy.inra.fr/unites/miaj/public/logiciels/R2Cuba/welcome.html`
  R interface to Thomas Hahn's `Cuba` library (see above) for deterministic and Monte Carlo cubature
- **rpy2** `http://rpy.sourceforge.net/rpy2.html`
  Provides access to R from Python; see also **PypeR** (`http://www.webarray.org/softwares/PypeR/`) for an alternative interface relying on pipes, with simpler installation requirements but less efficiency

C/C++/Fortran

- **BayeSys 3** `http://www.inference.phy.cam.ac.uk/bayesys/`
  Sophisticated suite of MCMC samplers including transdimensional capability, by the author of MemSys
- **fbm** `http://www.cs.utoronto.ca/~radford/fbm.software.html`
  Flexible Bayesian Modeling: MCMC for simple Bayes, nonparametric Bayesian regression and classification models based on neural networks and Gaussian processes, and Bayesian density estimation and clustering using mixture models and Dirichlet diffusion trees
- **BayesPack, DCUHRE**
  `http://www.sci.wsu.edu/math/faculty/genz/homepage`
  Adaptive quadrature, randomized quadrature, Monte Carlo integration
- **CUDAHM** Forthcoming C++ framework for accelerating hierarchical Bayesian methods (by astronomers Brandon Kelly, Tamas Budavari, TL)
- **BIE**, **CDF Bayesian limits**, **CUBA** (see above)

Java

- **Hydra** `http://research.warnes.net/projects/mcmc/hydra/`
  HYDRA provides methods for implementing MCMC samplers using Metropolis, Metropolis-Hastings, Gibbs methods. In addition, it provides classes implementing several unique adaptive and multiple chain/parallel MCMC methods.
- **YADAS** `http://www.stat.lanl.gov/yadas/home.html`
  Software system for statistical analysis using MCMC, based on the multi-parameter Metropolis-Hastings algorithm (rather than parameter-at-a-time Gibbs sampling)
- **Omega-hat** `http://www.omegahat.org/`
  Java environment for statistical computing, being developed by XLisp-stat and R developers

Other Statisticians' & Engineers' Tools

- **Stan** `http://mc-stan.org/`
  Budding successor to BUGS/JAGS, with a similar modeling language based on describing a generative model via conditional distributions for parameters and data; compiles models to C++; uses Hamiltonian Monte Carlo for posterior sampling, supported by automatic differentiation of models
- **JAGS** `http://www-fis.iarc.fr/~martyn/software/jags/`
  "Just Another Gibbs Sampler;" MCMC, esp. for Bayesian hierarchical models
- **BUGS/WinBUGS** `http://www.mrc-bsu.cam.ac.uk/bugs/`
  Bayesian Inference Using Gibbs Sampling: Flexible software for the Bayesian analysis of complex statistical models using MCMC
- **OpenBUGS** `http://mathstat.helsinki.fi/openbugs/`
  BUGS on Windows and Linux, and from inside the R
- **XLisp-stat** `http://www.stat.uiowa.edu/~luke/xls/xlsinfo/xlsinfo.html`
  Lisp-based data analysis environment, with an emphasis on providing a framework for exploring the use of dynamic graphical methods
- **ReBEL** `http://choosh.csee.ogi.edu/rebel/`
  Library supporting recursive Bayesian estimation in Matlab (Kalman filter, particle filters, sequential Monte Carlo).