ICTP Cosmology Summer School 2022

Statistics course: annotated references

Prof. Roberto Trotta, July 2022

Foundations of Bayesianism

Background reading on probability theory as reasoning under uncertainty (see chapter 2 for Jaynes' version of Cox Theorem):

E. T. Jaynes, Probability Theory: The Logic of Science. CUP (2003).

Introduction to Bayesian methods:

R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology. Contemp. Phys. **49**, 71-104 (2008).

and

R. Trotta, Bayesian Methods in Cosmology, (2017), Lecture notes for the 44th Saas Fee Advanced Course on Astronomy and Astrophysics, "Cosmology with wide-field surveys".

Historical background and importance of Bayes theorem (at a general audience level):

S. B. McGrayne, The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy, Tantor Media (2012)

Bias in coin tossing and probability as epistemic uncertainty:

P. Diaconis, S. Holmes, R. Montgomery, Dynamical Bias in the Coin Toss. *SIAM Review* **49**, 211-235 (2007).

Rigorous proof of Cox Theorem (see references therein for original papers):

K. S. Van Horn, Constructing a logic of plausible inference: a guide to Cox's theorem. *International Journal of Approximate Reasoning* **34**, 3-24 (2003).

Philosophical and historical aspects:

M.C. Galavotti, Philosophical Introduction to Probability, CSLI Publications (2009)

Statement, proof and consequences of de Finetti's Representation Theorem (in its original form, a result he obtained in 1937):

J. M. Bernardo, The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences* **4**, 111-122 (1996).

A. P. Dawid, Probability, Causality and the Empirical World: A Bayes-de Finetti-Popper- Borel Synthesis. *Statist. Sci.* **19**, 44-57 (2004).

D. J. Poirier, Excheangeability, Representation Theorems and Subjectivity. In: The Oxford Handbook of Bayesian Econometrics, J. Geweke, G. Koop, & H. Van Dijk (Eds), OUP (2010).

Review article on prior selection:
R. E. Kass, L. Wasserman, The Selection of Prior Distributions by Formal Rules. Journal of the American Statistical Association 91, 1343–1370 (1996).

Reference priors (as in "maximal learning priors" or "objective priors"):
J. O. Berger, J. M. Bernardo, D. Sun, The formal definition of reference priors. *Ann. Statist.* **37**, 905-938 (2009).
J. M. Bernardo, Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society: Series B (Methodological)* **41**, 113-128 (1979).
J. M. Bernardo, "Reference Analysis" in Handbook of Statistics, C. R. R. D. K. Dey, Ed. (Elsevier, 2005), vol. 25, pp. 17-90.
J.M. Bernardo, Bayesian Reference Analysis. A Postgraduate Tutorial Course online, (1998).

Proof that Jeffreys's prior maximises Shannon's mutual information asymptotically (conjectured by Bernardo, 1979):
B. S. Clarke, A. R. Barron, Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference* **41**, 37-60 (1994).

Empirical Bayes and James-Stein estimator:
B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Institute of Mathematical Statistics Monographs (Cambridge University Press, Cambridge, 2010), DOI: 10.1017/CBO9780511761362 (chapter 1)

B. Efron, C. Morris, Stein's Paradox in Statistics. *Scientific American - SCI AMER* **236**, 119-127 (1977).

Sampling Methods

General reference:
A. Gelman *et al.*, Bayesian Data Analysis, Chapman Hall/CRC (3rd ed, 2013).

MCMC convergence diagnostics:
S. P. Brooks, A. Gelman, General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* **7**, 434-455 (1998).
S. P. Brooks, G. O. Roberts, Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing* **8**, 319-335 (1998).
M. K. Cowles, B. P. Carlin, Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association* **91**, 883-904 (1996).
S. Rosenthal, A review of asymptotic convergence for general state space Markov chains. *Far East Journal of Theoretical Statistics* **5** (2001).
V. Roy, Convergence diagnostics for Markov chain Monte Carlo, preprint: 1909.11827.

Optimal choice of jumping kernel for MH:
A. Gelman, G. O. Roberts, W. R. Gilks, Efficient Metropolis jumping rules. *Bayesian statistics* **5**, 42 (1996).

A. Gelman, D. B. Rubin, Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* **7**, 457-472 (1992).

D. Vats, C. Knudson, Revisiting the Gelman-Rubin Diagnostic, preprint: 1812.09384

Adaptive Metropolis algorithm (and proof of convergence):

H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm. *Bernoulli* **7**, 223-242 (2001).

Gibbs sampling:

A. F. M. Smith, G. O. Roberts, Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Methodological)* **55**, 3-23 (1993).

S. Geman, D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721-741 (1984).

Hamiltonian Monte Carlo

R. M. Neal (2011), MCMC using Hamiltonian dynamics, in: "Handbook of Markov Chain Monte Carlo" (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, Eds), Chapman & Hall/CRC Press.

M. Betancourt (2017), A Conceptual Introduction to Hamiltonian Monte Carlo, arXiv: 1701.02434

M. D. Homan, A. Gelman, The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).

Bayesian Model comparison (BMC)

The p-value fallacy:

S. Greenland *et al.*, Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* **31**, 337-350 (2016).

J. Gill, The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* **52**, 647-674 (1999).

S. N. Goodman, Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. *Annals of Internal Medicine* **130**, 995-1004 (1999).

R. Hubbard, M. J. Bayarri, Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing. *The American Statistician* **57**, 171-178 (2003).

B. Vidgen, T. Yasseri, P-Values: Misunderstood and Misused. *Frontiers in Physics* **4** (2016).

R. L. Wasserstein, N. A. Lazar, The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician* **70**, 129-133 (2016).

Difference/calibration between frequentist hypothesis test and BMC:

J. O. Berger, Could Fisher, Jeffreys and Neyman Have Agreed on Testing? *Statist. Sci.* **18**, 1-32 (2003).

J. O. Berger, T. Sellke, Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association* **82**, 112-122 (1987).

S. N. Goodman, Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of internal medicine* **130**, 1005-1013 (1999).

E. L. Lehmann, The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? *Journal of the American Statistical Association* **88**, 1242-1249 (1993).
T. Sellke, M. J. Bayarri, J. O. Berger, Calibration of ρ Values for Testing Precise Null Hypotheses. *The American Statistician* **55**, 62-71 (2001).
I. J. Good (1950) Probability and the Weighing of Evidence. (C. Griffin London).

Methods for prior choice for BMC:

J. O. Berger, L. R. Pericchi, The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association* **91**, 109-122 (1996).
J. O. Berger *et al.*, Objective Bayesian methods for model selection: Introduction and comparison. *Lecture Notes-Monograph Series*, 135-207 (2001).
G. Consonni, D. Fouskakis, B. Liseo, I. Ntzoufras, Prior Distributions for Objective Bayesian Analysis. *Bayesian Analysis* **13**, 627-679, 653 (2018).
L. Wasserman, Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology* **44**, 92-107 (2000).

Jeffreys-Lindley paradox:

R. D. Cousins (2013), The Jeffreys-Lindley Paradox and Discovery Criteria in High Energy Physics., arxiv:1310:3791
L. Demortier and L. Lyons, Testing Hypotheses in Particle Physics: Plots of p0 versus p1, technical report (2014).

Computational methods:

T. J. Diciccio, R. E. Kass, A. Raftery, L. Wasserman, Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association* **92**, 903-915 (1997).
F. Feroz, M. P. Hobson, Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society* **384**, 449 (2008).
A. Heavens *et al.* (2017), Marginal Likelihoods from Monte Carlo Markov Chains, arxiv: 1704.03472
A. A. Neath, J. E. Cavanaugh, The Bayesian information criterion: background, derivation, and applications. *WIREs Computational Statistics* **4**, 199-203 (2012).
J. Skilling, Nested Sampling. *AIP Conference Proceedings* **735**, 395-405 (2004).
J. Skilling, Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**, 833-859, 827 (2006).
I. Verdinelli, L. Wasserman, Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association* **90**, 614-618 (1995).
G. Albert (2020), JAXNS: a high-performance nested sampling package based on JAX, arxiv: 2012.15286