

Prior Probabilities: A Brief Bibliography

Wray L. Buntine

Ultimode Systems

555 Bryant St. #186

Palo Alto, CA, 94301, USA

wray@ultimode.com

Phone: (415) 324 3447

A broad discussion of priors can be found in (Bernardo & Smith, 1994; Kass & Wasserman, 1996) including objectivity, reference priors, priors for some specific distributions, a number of paradoxes, and connections with cross validation. Good examples of methods can be found in (Gelman, Carlin, Stern, & Rubin, 1995). Another source of general discussion is papers by Jaynes collected in (Rosenkrantz, 1983). Ockham's razor is discussed from several perspectives in (Barron & Cover, 1991; Jefferys & Berger, 1992; Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987; Cover & Thomas, 1991; Wallace & Freeman, 1987). Objectivity is also discussed in (Berger, 1988, 1985). General principles for priors are discussed in (Bernardo & Smith, 1994; Kass & Wasserman, 1996). Paradoxes are discussed in (Jaynes, 1980, 1973; Diaconis & Freedman, 1986; Bernardo & Smith, 1994). An introduction to eliciting probabilities is given in (Morgan & Henrion, 1990; Heckerman & Jimison, 1989). The use of measure theory for priors is given in (Wootters, 1981). Overfitting or the bias-variance dilemma is discussed in (Geman, Bienenstock, & Dourson, 1992).

Standard priors for specific distributions, and conjugate priors are discussed in (Box & Tiao, 1973; Bernardo & Smith, 1994; DeGroot, 1970). Priors for feed-forward neural networks, an extensive research topic, are discussed in (Nowlan & Hinton, 1992; MacKay, 1992; Buntine & Weigend, 1991; Wolpert, 1994; Neal, 1996, 1994; MacKay, 1995). Priors for Bayesian networks are discussed in (Heckerman, Geiger, & Chickering, 1994; Buntine, 1991b; Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993). Priors for class probability trees are discussed in (Buntine, 1991a; Wallace & Patrick, 1993). General methods for dealing with priors, empirical Bayes, hierarchical priors, ML-II priors, etc., are discussed in (Bernardo & Smith, 1994; Berger, 1985).

References

- Barron, A., & Cover, T. (1991). Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37(4).
- Berger, J. O. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, 76(March-April), 159–165.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- Bernardo, J., & Smith, A. (1994). *Bayesian Theory*. John Wiley, Chichester.
- Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. (1987). Occam's razor. *Information Processing Letters*, 24, 377–380.
- Box, G., & Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts.
- Buntine, W. (1991a). Learning classification trees. In Hand, D. (Ed.), *Artificial Intelligence Frontiers in Statistics*, pp. 182–201. Chapman & Hall, London.
- Buntine, W. (1991b). Theory refinement of Bayesian networks. In D'Ambrosio, B., Smets, P., & Bonissone, P. (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Seventh Conference* Los Angeles, CA.
- Buntine, W., & Weigend, A. (1991). Bayesian back-propagation. *Complex Systems*, 5(1), 603–643.
- Cover, T., & Thomas, J. (1991). *Elements of Information Theory*. John Wiley & Sons.
- DeGroot, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill.

- Diaconis, P., & Freedman, D. (1986). On inconsistent Bayes estimates of location. *Annals of Statistics*, 14(1), 68–87.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (1995). *Bayesian Data Analysis*. Chapman & Hall.
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1–58.
- Heckerman, D., Geiger, D., & Chickering, D. (1994). Learning Bayesian networks: The combination of knowledge and statistical data. Technical report MSR-TR-94-09 (Revised), Microsoft Research, Advanced Technology Division. To appear, *Machine Learning Journal*.
- Heckerman, D., & Jimison, H. (1989). A perspective on confidence and its use in focusing attention during knowledge acquisition. In Kanal, L. N., Levitt, T. S., & Lemmer, J. F. (Eds.), *Uncertainty in Artificial Intelligence 3*, pp. 123–131 Amsterdam. Elsevier Science Publishers.
- Jaynes, E. (1973). The well-posed problem. *Foundations of Physics*, 3, 477–493.
- Jaynes, E. (1980). Marginalization and prior probabilities. In Zellner, A. (Ed.), *Bayesian Analysis in Econometrics and Statistics*. North-Holland, Amsterdam.
- Jefferys, W., & Berger, J. (1992). Ockham’s razor and Bayesian analysis. *American Scientist*, 80(Jan-Feb), 64–72.
- Kass, R., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1343–1370.
- MacKay, D. (1992). Bayesian interpolation. *Neural Computation*, 4(3), 415–447.
- MacKay, D. (1995). Probable networks and plausible predictions - a review of practical bayesian methods for supervised neural networks. *Network, IOPP*. Submitted.
- Morgan, M., & Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.
- Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag, New York. Revision of PhD Thesis from University of Toronto, abstract at <http://www.cs.utoronto.ca/~radford/bnn.book.html>.
- Neal, R. (1994). Priors for infinite networks. Technical report CRG-TR-94-1, Dept. of Computer Science, University of Toronto.
- Nowlan, S., & Hinton, G. (1992). Simplifying neural networks by soft weight sharing. In Touretzky, D. (Ed.), *Advances in Neural Information Processing Systems 4 (NIPS*91)*. Morgan Kaufmann.
- Rosenkrantz, R. (Ed.). (1983). *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. D. Reidel Publishing.
- Spiegelhalter, D., Dawid, A., Lauritzen, S., & Cowell, R. (1993). Bayesian analysis in expert systems. *Statistical Science*, 8(3), 219–283.
- Wallace, C., & Freeman, P. (1987). Estimation and inference by compact encoding. *Journal of the Royal Statistical Society B*, 49(3), 240–265.
- Wallace, C., & Patrick, J. (1993). Coding decision trees. *Machine Learning*, 11, 7–22.
- Wolpert, D. (1994). Bayesian backpropagation over functions rather than weights. In Tesauro, G. (Ed.), *Advances in Neural Information Processing Systems 6 (NIPS*93)*. Morgan Kaufmann.
- Wootters, W. (1981). Statistical distance and Hilbert space. *Physical Review D*, 23(2).