

Bibliography

- D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–169, 1985. [2.4.1](#)
- N. J. Adams, A. J. Storkey, Z. Ghahramani, and C. K. I. Williams. MFDTs: Mean field dynamic trees. In *Proc. 15th Int. Conf. on Pattern Recognition*, 2000. [2.5.1](#)
- S. L. Adler. Over-relaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions. *Physical Review D*, 23:2901–2904, 1981. [1.3.6](#)
- H. Attias. Independent Factor Analysis. *Neural Computation*, 11:803–851, 1999a. [2.4.1](#)
- H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *Proc. 15th Conf. on Uncertainty in Artificial Intelligence*, 1999b. [2.6.1](#)
- H. Attias. A variational Bayesian framework for graphical models. In S. A. Solla, T. K. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems 12*, Cambridge, MA, 2000. MIT Press. [2.3.2](#), [2.4.3](#)
- L. Bahl and F. Jelinek. Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition. *IEEE Transactions on Information Theory*, 21(4):404–411, 1975. [3.1](#)
- D. Barber and C. M. Bishop. Ensemble learning for multi-layer networks. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 395–401, Cambridge, MA, 1998. MIT Press. [2.3.2](#)
- D. Barber and P. Sollich. Gaussian fields for approximate inference in layered sigmoid belief networks. In S. A. Solla, T. K. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems 12*, Cambridge, MA, 2000. MIT Press. [2.3.2](#)
- A. I. Barvinok. Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor. *Random Structures and Algorithms*, 14(1):29–61, 1999. [1.3.3](#)
- L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37(6):1554–1563, 1966. [3.1](#)

- L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41:164–171, 1970. 2.2.2, 3.1, 3.2
- M. J. Beal, Z. Ghahramani, and C. E. Rasmussen. The infinite hidden Markov model. In *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2002. MIT Press. 3.6, 7.1
- A. J. Bell and T. J. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995. 2.4.1
- J. M. Bernardo and F. J. Giron. A Bayesian analysis of simple mixture problems. In J. M. Bernardo, M. H. Degroot, A. F. Smith, and D. V. Lindley, editors, *Bayesian Statistics 3*, pages 67–78. Clarendon Press, 1988. 2.3.2
- J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, Inc., New York, 1994. 1.2.2, 2.6.1
- C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995. 2.4.1
- C. M. Bishop. Variational PCA. In *Proc. Ninth Int. Conf. on Artificial Neural Networks. ICANN*, 1999. 2.3.2, 4.2.2
- C. M. Bishop, N. D. Lawrence, T. S. Jaakkola, and M. I. Jordan. Approximating posterior distributions in belief networks using mixtures. In *Advances in Neural Information Processing Systems 10*, Cambridge, MA, 1998. MIT Press. 2.3.2
- C. M. Bishop, D. Spiegelhalter, and J. Winn. VIBES: A variational inference engine for Bayesian networks. In *Advances in Neural Information Processing Systems 15*, S. Becker and S. Thrun and K. Obermayer, 2003. 2.4.3, 7.1
- X. Boyen and D. Koller. Tractable inference for complex stochastic processes. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, Wisconsin, 1998. 2.3.2
- S. Brooks. MCMC repository. Statistical Laboratory, University of Cambridge. Accessible on the world wide web at <http://www.statslab.cam.ac.uk/~mcmc>. 1.3.6
- W. Buntine. Variational extensions to EM and multinomial PCA. In *ECML*, 2002. 2.4.3
- G. Casella, K. L. Mengersen, C. P. Robert, and D. M. Titterton. Perfect slice samplers for mixtures of distributions. *Journal of the Royal Statistical Society, Series B (Methodological)*, 64(4):777–790, 2000. 1.3.6
- K. Chan, T. Lee, and T. J. Sejnowski. Variational learning of clusters of undercomplete nonsymmetric independent components. *Journal of Machine Learning Research*, 3:99–114, August 2002. 4.8

- P. Cheeseman and J. Stutz. Bayesian classification (Autoclass): Theory and results. In U. M. Fayyad, G. Piatesky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 153–180, Menlo Park, CA, 1996. AAAI Press/MIT Press. 1.3.1, 1.3.5, 1.3.5, 2.6.2, 2.6.2, 6.3.3, 6.5.2
- D. M. Chickering and D. Heckerman. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 29(2–3):181–212, 1997. 2.6.2, 2.6.2
- R. Choudrey and S. Roberts. Variational mixture of Bayesian independent component analysers. *Neural Computation*, 15(1), 2002. 4.8
- P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994. 2.4.1
- R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York, 1999. 1.1
- R. T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946. 1.1
- N. de Freitas, P. Højten-Sørensen, M. I. Jordan, and S. Russell. Variational MCMC. In J. S. Breese and D. Koller, editors, *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2001. 4.8
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39:1–38, 1977. 2.2.2, 2.4.3
- R. P. Feynman. *Statistical Mechanics: A Set of Lectures*. Perseus, Reading, MA, 1972. 2.2.1, 2.3.2
- J. A. Fill. An interruptible algorithm for perfect sampling via Markov chains. *The Annals of Applied Probability*, 8(1):131–162, 1998. 1.3.6
- E. Fokoué and D. M. Titterton. Mixtures of factor analysers. Bayesian estimation and inference by stochastic simulation. *Machine Learning*, 50(1):73–94, January 2003. 4.2.3
- B. J. Frey, R. Patrascu, T. S. Jaakkola, and J. Moran. Sequentially fitting “inclusive” trees for inference in noisy-OR networks. In *Advances in Neural Information Processing Systems 13*, 2001. 2.3.2
- N. Friedman. The Bayesian structural EM algorithm. In *Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI '98, San Francisco, CA, 1998)*. Morgan Kaufmann Publishers. 6.4, 6.6

- N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7:601–620, 2000. 5.5
- S. Frühwirth-Schnatter. Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57: 237–246, 1995. 5.6
- D. Geiger, D. Heckerman, and C. Meek. Asymptotic model selection for directed networks with hidden variables. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, 1996. 6.5.2
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 1995. 2.4.1
- A. Gelman and X. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13:163–185, 1998. 6.3.5
- Z. Ghahramani. Factorial learning and the EM algorithm. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 617–624, Cambridge, MA, 1995. MIT Press. 2.2.3
- Z. Ghahramani. An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001. 3.1
- Z. Ghahramani and H. Attias. Online variational Bayesian learning, 2000. Slides from talk presented at NIPS 2000 workshop on Online Learning, available at <http://www.gatsby.ucl.ac.uk/~zoubin/papers/nips00w.ps>. 2.4.3
- Z. Ghahramani and M. J. Beal. Variational inference for Bayesian mixtures of factor analyzers. In *Advances in Neural Information Processing Systems 12*, Cambridge, MA, 2000. MIT Press. 2.3.2, 2.4.3, 4.1, 4.7
- Z. Ghahramani and M. J. Beal. Propagation algorithms for variational Bayesian learning. In *Advances in Neural Information Processing Systems 13*, Cambridge, MA, 2001. MIT Press. 2.4.3, 2.4.3
- Z. Ghahramani and G. E. Hinton. Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, Department of Computer Science, University of Toronto, 1996a. 5.2.2, 5.3.8
- Z. Ghahramani and G. E. Hinton. The EM algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996b. 4.1.2
- Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. *Neural Computation*, 12(4), 2000. 2.2.3, 3.1

- Z. Ghahramani and M. I. Jordan. Factorial hidden Markov models. *Machine Learning*, 29: 245–273, 1997. 2.2.3, 2.2.3, 3.1
- W. R. Gilks. Derivative-free adaptive rejection sampling for Gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 641–649. Clarendon Press, 1992. 1.3.6
- W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 44:455–472, 1995. 1.3.6
- W. R. Gilks, G. O. Roberts, and S. K. Sahu. Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93:1045–1054, 1998. 4.8
- W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2):337–348, 1992. 1.3.6, 1.3.6
- A. G. Gray, B. Fischer, J. Schumann, and W. Buntine. Automatic derivation of statistical algorithms: The EM family and beyond. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003. 2.4.3, 7.1
- P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995. 1.3.6, 4.2.3, 4.3
- M. Harvey and R. M. Neal. Inference for belief networks using coupling from the past. In C. Boutilier and M. Goldszmidt, editors, *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 256–263. Morgan Kaufmann, 2000. 1.3.6
- W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. 1.3.6, 6.3.5
- D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06 [ftp://ftp.research.microsoft.com/pub/tr/TR-95-06.PS] , Microsoft Research, 1996. 1.1
- D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995. 6.2, 7.1
- T. Heskes. Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, Cambridge, MA, 2003. MIT Press. 7.1
- G. E. Hinton and D. van Camp. Keeping neural networks simple by minimizing the description length of the weights. In *Sixth ACM Conference on Computational Learning Theory*, Santa Cruz, 1993. 1.2.1, 2.3.2
- G. E. Hinton and R. S. Zemel. Autoencoders, minimum description length, and Helmholtz free energy. In J. D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*, San Francisco, CA, 1994. Morgan Kaufmann. 2.2.3

- J. J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), May 1994. 4.6
- T. S. Jaakkola. Variational methods for inference and estimation in graphical models. Technical Report Ph.D. Thesis, Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, 1997. 2.2.3, 2.3.2, 2.5.2, 7.1
- T. S. Jaakkola and M. I. Jordan. Improving the mean field approximation via the use of mixture distributions. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 163–173. Kluwer, 1998. 2.3.2
- T. S. Jaakkola and M. I. Jordan. Bayesian logistic regression: a variational approach. *Statistics and Computing*, 10:25–37, 2000. 2.3.2
- E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. 1.1
- W. H. Jefferys and J. O. Berger. Ockham’s razor and Bayesian analysis. *American Scientist*, 80: 64–72, 1992. 1.2.1, 4.2
- H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007), 1946. 1.2.2
- F. V. Jensen. *Introduction to Bayesian Networks*. Springer-Verlag, New York, 1996. 1.1
- J. L. W. V. Jensen. Sur les fonctions convexes et les inegalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906. 2.2.1
- M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. In *ACM Symposium on Theory of Computing*, pages 712–721, 2001. 1.3.3
- M. I. Jordan, editor. *Learning in Graphical Models*. MIT Press, Cambridge, MA, 1999. 1.1
- M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods in graphical models. *Machine Learning*, 37:183–233, 1999. 2.3.2, 2.5.2
- M. I. Jordan, Z. Ghahramani, and L. K. Saul. Hidden Markov decision trees. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, Cambridge, MA, 1997. MIT Press. 3.1
- M. I. Jordan and Y. Weiss. Graphical models: Probabilistic inference. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks, 2nd edition*. MIT Press, Cambridge, MA, 2002. 1.1.2
- B. H. Juang and L. R. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991. 3.1

- R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995. 1.2.1, 1.3.1, 1.3.2
- S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. 6.3.5
- T. Kočka and N. L. Zhang. Dimension correction for hierarchical latent class models. In A. Darwiche and N. Friedman, editors, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 267–274. Morgan Kaufmann, 2002. 6.5.2
- S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 50(2):157–224, 1988. 1.1
- N. D. Lawrence and M. Azzouzi. A variational Bayesian committee of neural networks. Submitted to *Neural Networks*, 1999. 2.3.2
- N. D. Lawrence, C. M. Bishop, and M. I. Jordan. Mixture representations for inference and learning in Boltzmann machines. In G. F. Cooper and S. Moral, editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 320–327, Madison, Wisconsin, 1998. 2.3.2
- Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995. 4.6.2
- M. A. R. Leisink and H. J. Kappen. A tighter bound for graphical models. *Neural Computation*, 13(9):2149–2170, 2001. 7.1
- D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992. 3.3, 4.2
- D. J. C. MacKay. Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6:469–505, 1995. 1.2.1, 1.3, 1.3.2, 5.2.2, 6.3.2
- D. J. C. MacKay. Bayesian non-linear modelling for the 1993 energy prediction competition. In G. Heidbreder, editor, *Maximum Entropy and Bayesian Methods, Santa Barbara 1993*, pages 221–234, Dordrecht, 1996. Kluwer. 4.2.2
- D. J. C. MacKay. Ensemble learning for hidden Markov models. Technical report, Cavendish Laboratory, University of Cambridge, 1997. 2.3.2, 2.4.3, 2.5.1, 3.1, 3.4, 3.4.2, 3.5.2, 7.2
- D. J. C. MacKay. Choice of basis for Laplace approximation. *Machine Learning*, 33(1), 1998. 1.3.2, 3.3, 3.3, 6.3.1
- D. J. C. MacKay. An introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*. MIT Press, Cambridge, MA, 1999. 4.7.3

- D. J. C. MacKay. A problem with variational free energy minimization, 2001. 3.6
- D. J. C. MacKay and L. C. Peto. A hierarchical Dirichlet language model. *Natural Language Engineering*, 1(3):1–19, 1995. 3.3
- N. Metropolis, A. W. Rosenbluth, M. N. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953. 1.3.6, 6.3.5
- T. P. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, MIT, 2001a. 2.3.2, 7.1
- T. P. Minka. Using lower bounds to approximate integrals, 2001b. 2.6.2, 7.2
- T. P. Minka and J. Lafferty. Expectation-Propagation for the generative aspect model. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2002. 2.3.2
- J. W. Miskin. *Ensemble Learning for Independent Component Analysis*. PhD thesis, University of Cambridge, December 2000. 4.7, 4.7.3, 5.6
- D. J. Murdoch and P. J. Green. Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25(3):483–502, 1998. 1.3.6
- R. M. Neal. Connectionist learning of belief networks. *Artificial Intelligence*, 56:71–113, 1992. 1.3.6, 2.2.3
- R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993. 1.3.6, 6.3.5
- R. M. Neal. *Bayesian Learning in Neural Networks*. Springer-Verlag, 1996. 1.2.1, 1.3.6, 3.6, 6.3.5, 7.1
- R. M. Neal. Assessing relevance determination methods using DELVE. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, pages 97–129. Springer-Verlag, 1998a. 4.2.2
- R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. Technical Report 9815, Department of Statistics, University of Toronto, 1998b. 3.6
- R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11:125–139, 2001. 1.3.6, 6.3.5, 6.3.5, 6.3.5
- R. M. Neal. Slice sampling. *Annals of Statistics*, 31(3), 2003. With discussion. 6.6
- R. M. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–369. Kluwer Academic Publishers, 1998. 2.2.1, 2.3.2

- A. O'Hagan. Monte Carlo is fundamentally unsound. *Statistician*, 36(2/3):247–249, 1987. Special Issue: Practical Bayesian Statistics. 1.3.6
- A. O'Hagan. Bayes-Hermite quadrature. *Journal of Statistical Planning and Inference*, 29(3): 245–260, 1991. 1.3.6
- G. Parisi. *Statistical Field Theory*. Addison Wesley, 1988. 2.3.2
- J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Francisco, CA, 1988. 1.1, 1.1.1, 2.5
- J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 1&2(9):223–252, 1996. 1.3.6
- L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE Acoustics, Speech & Signal Processing Magazine*, 3:4–16, 1986. 3.2
- C. Rangel, D. L. Wild, F. Falciani, Z. Ghahramani, and A. Gaiba. Modeling biological responses using gene expression profiling and linear dynamical systems. In *To appear in Proceedings of the 2nd International Conference on Systems Biology*, Madison, WI, 2001. OmniPress. 5.5, 5.9
- C. E. Rasmussen. The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems 12*, Cambridge, MA, 2000. MIT Press. 3.6, 4.8, 7.1
- C. E. Rasmussen and Z. Ghahramani. Occam's razor. In *Advances in Neural Information Processing Systems 13*, Cambridge, MA, 2001. MIT Press. 1.2.1
- C. E. Rasmussen and Z. Ghahramani. Infinite mixtures of Gaussian process experts. In *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2002. MIT Press. 7.1
- C. E. Rasmussen and Z. Ghahramani. Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2003. MIT Press. 1.3.6
- H. E. Rauch. Solutions to the linear smoothing problem. *IEEE Transactions on Automatic Control*, 8:371–372, 1963. 5.3.4
- H. E. Rauch, F. Tung, and C. T. Striebel. On the maximum likelihood estimates for linear dynamic systems. Technical Report 6-90-63-62, Lockheed Missiles and Space Co., Palo Alto, California, June 1963. 5.3.2
- S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B (Methodological)*, 59(4): 731–758, 1997. 4.3
- J. Rissanen. Stochastic complexity. *Journal of the Royal Statistical Society, Series B (Methodological)*, 49:223–239 and 253–265, 1987. With discussion. 1.3.4

- C. P. Robert, G. Celeux, and J. Diebolt. Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statistics & Probability Letters*, 16(1):77–83, 1993. 3.3
- S. J. Roberts, D. Husmeier, I. Rezek, and W. Penny. Bayesian approaches to Gaussian mixture modeling. *IEEE PAMI*, 20(11):1133–1142, 1998. 4.2.3
- S. T. Roweis and Z. Ghahramani. A unifying review of linear Gaussian models. *Neural Computation*, 11(2):305–345, 1999. 4.1.2, 5.2.1
- J. Särelä, H. Valpola, R. Vigário, and E. Oja. Dynamical factor analysis of rhythmic magnetoencephalographic activity. In *Proceedings of the 3rd International Conference on Independent Component Analysis and Blind Signal Separation, ICA 2001*, pages 457–462, San Diego, California, USA, 2001. 5.6
- M. Sato. Online model selection based on the variational Bayes. *Neural Computation*, 13(7):1649–1681, 2001. 2.4.3
- L. K. Saul, T. Jaakkola, and M. I. Jordan. Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, 4:61–76, 1996. 2.2.3
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978. 1.3.1, 1.3.4
- R. Settini and J. Q. Smith. On the geometry of Bayesian graphical models with hidden variables. In G. F. Cooper and S. Moral, editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 472–479, Madison, Wisconsin, 1998. 6.5.2
- R. D. Shachter. Bayes-Ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In G. F. Cooper and S. Moral, editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 480–487, Madison, Wisconsin, 1998. Morgan Kaufmann. 1.1.1
- R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264, 1982. 5.2.2
- P. Smyth, D. Heckerman, and M. I. Jordan. Probabilistic independence networks for hidden Markov probability models. *Neural Computation*, 9:227–269, 1997. 3.1
- M. Stephens. *Bayesian methods for mixtures of normal distributions*. PhD thesis, Oxford University, 1997. 2.3.2
- A. Stolcke and S. Omohundro. Hidden Markov model induction by Bayesian model merging. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems 5*, pages 11–18, San Francisco, CA, 1993. Morgan Kaufmann. 3.3

- A. M. Storkey. Dynamic trees: A structured variational method giving efficient propagation rules. In *Proc. 16th Conf. on Uncertainty in Artificial Intelligence. UAI*, San Francisco, CA, 2000. Morgan Kaufmann Publishers. 2.5.1
- A. Thomas, D. J. Spiegelhalter, and W. R. Gilks. BUGS: A program to perform Bayesian inference using Gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 837–842. Clarendon Press, 1992. 7.1
- M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999. 4.1.2
- G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. *J. Comp. Phys.*, 23:187–199, 1977. 6.3.5
- N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. SMEM algorithm for mixture models. *Neural Computation*, 12(9):2109–2128, 2000. 4.3, 4.3.2, 4.5.3, 4.5, 4.6.2
- H. Valpola and J. Karhunen. An unsupervised ensemble learning method for nonlinear dynamic state-space models. *Neural Computation*, 14(11):2647–2692, 2002. 5.6
- A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967. 3.3
- M. J. Wainwright, T. Jaakkola, and A. S. Willsky. A new class of upper bounds on the log partition function. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 2002. 7.1
- C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society, Series B (Methodological)*, 49(3):240–265, 1987. With discussion. 1.3.4
- S. Waterhouse, D. J. C. MacKay, and T. Robinson. Bayesian methods for mixtures of experts. In *Advances in Neural Information Processing Systems 8*, Cambridge, MA, 1996. MIT Press. 2.3.2
- M. Welling and Y. W. Teh. Belief Optimisation for binary networks: A stable alternative to loopy belief propagation. In *UAI 2001*, Seattle, Washington, 2001. 7.1
- C. K. I. Williams and N. J. Adams. DTs: Dynamic trees. In *Advances in Neural Information Processing Systems 11*, Cambridge, MA, 1999. MIT Press. 2.5.1
- C. K. I. Williams and G. E. Hinton. Mean field networks that learn to discriminate temporally distorted strings. In D. S. Touretzky, J. L. Elman, T. J. Sejnowski, and G. E. Hinton, editors, *Connectionist Models: Proceedings of the 1990 Summer School*, pages 18–22. Morgan Kaufmann Publishers, San Francisco, CA, 1991. 2.2.3

- J. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, Cambridge, MA, 2001. MIT Press. 7.1
- A. L. Yuille. CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation. Technical report, Smith-Kettlewell Eye Research Institute, 2001. 7.1