Bibliography

Abbeel, P., D. Koller, and A. Ng (2006, August). Learning factor graphs in polynomial time & sample complexity. *Journal of Machine Learning Research 7*, 1743–1788.

Ackley, D., G. Hinton, and T. Sejnowski (1985). A learning algorithm for Boltzmann machines. *Cognitive Science 9*, 147–169.

Aji, S. M. and R. J. McEliece (2000). The generalized distributive law. *IEEE Trans. Information Theory 46*, 325–343.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control 19*, 716–723.

Akashi, H. and H. Kumamoto (1977). Random sampling approach to state estimation in switching environments. *Automatica 13*, 429–434.

Allen, D. and A. Darwiche (2003a). New advances in inference by recursive conditioning. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 2–10.

Allen, D. and A. Darwiche (2003b). Optimal time–space tradeoff in probabilistic inference. In *Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 969–975.

Altun, Y., I. Tsochantaridis, and T. Hofmann (2003). Hidden Markov support vector machines. In *Proc. 20th International Conference on Machine Learning (ICML)*.

Andersen, S., K. Olesen, F. Jensen, and F. Jensen (1989). HUGIN—a shell for building Bayesian belief universes for expert systems. In *Proc. 11th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1080–1085.

Anderson, N. (1974). Information integration theory: A brief survey. In *Contemporary developments in Mathematical Psychology*, Volume 2, pp. 236–305. San Francisco, California: W.H. Freeman and Company.

Anderson, N. (1976). How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology 61*(6), 677–692.

Andreassen, S., F. Jensen, S. Andersen, B. Falck, U. Kjærulff, M. Woldbye, A. R. Sørensen, A. Rosenfalck, and F. Jensen (1989). MUNIN — an expert EMG assistant. In J. E. Desmedt (Ed.), *Computer-Aided Electromyography and Expert Systems*, Chapter 21. Amsterdam: Elsevier Science Publishers.

Anguelov, D., D. Koller, P. Srinivasan, S. Thrun, H.-C. Pang, and J. Davis (2004). The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *Proc. 18th Conference on Neural Information Processing Systems (NIPS)*.

Arnauld, A. and P. Nicole (1662). Port-royal logic.

Arnborg, S. (1985). Efficient algorithms for combinatorial problems on graphs with bounded, decomposability—a survey. *BIT 25*(1), 2–23.

Arnborg, S., D. Corneil, and A. Proskurowski (1987). Complexity of finding embeddings in a k-tree. *SIAM J. Algebraic Discrete Methods 8*(2), 277–284.

Attias, H. (1999). Inferring parameters and structure of latent variable models by variational Bayes. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 21–30.

Avriel, M. (2003). *Nonlinear Programming: Analysis and Methods*. Dover Publishing.

Bacchus, F. and A. Grove (1995). Graphical models for preference and utility. In *Proc. UAI–95*, pp. 3–10.

Bach, F. and M. Jordan (2001). Thin junction trees. In *Proc. 15th Conference on Neural Information Processing Systems (NIPS)*.

Balke, A. and J. Pearl (1994a). Counterfactual probabilities: Computational methods, bounds and applications. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 46–54.

Balke, A. and J. Pearl (1994b). Probabilistic evaluation of counterfactual queries. In *Proc. 10th Conference on Artificial Intelligence (AAAI)*, pp. 230–237.

Bar-Shalom, Y. (Ed.) (1992). *Multitarget multisensor tracking: Advanced applications*. Norwood, Massachusetts: Artech House.

Bar-Shalom, Y. and T. Fortmann (1988). *Tracking and Data Association*. New York: Academic Press.

Bar-Shalom, Y., X. Li, and T. Kirubarajan (2001). *Estimation with Application to Tracking and Navigation*. John Wiley and Sons.

Barash, Y. and N. Friedman (2002). Context-specific Bayesian clustering for gene expression data. *Journal of Computational Biology 9*, 169–191.

Barber, D. and W. Wiegerinck (1998). Tractable variational structures for approximating graphical models. In *Proc. 12th Conference on Neural Information Processing Systems (NIPS)*, pp. 183–189.

Barbu, A. and S. Zhu (2005). Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27*(8), 1239–1253.

Barnard, S. (1989). Stochastic stero matching over scale. *International Journal of Computer Vision 3*, 17–32.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley.

Barron, A., J. Rissanen, and B. Yu (1998). The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory 44*(6), 2743–2760.

Bartlett, M. (1935). Contingency table interactions. *Journal of the Royal Statistical Society, Series B 2*, 248–252.

Bauer, E., D. Koller, and Y. Singer (1997). Update rules for parameter estimation in Bayesian networks. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 3–13.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London 53*, 370–418.

Beal, M. and Z. Ghahramani (2006). Variational Bayesian learning of directed graphical models with hidden variables. *Bayesian Analysis 1*, 793–832.

Becker, A., R. Bar-Yehuda, and D. Geiger (1999). Random algorithms for the loop cutset problem. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 49–56.

Becker, A. and D. Geiger (1994). The loop cutset problem. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 60–68.

Becker, A. and D. Geiger (2001). A sufficiently fast algorithm for finding close to optimal clique

trees. *Artificial Intelligence 125*(1–2), 3–17.

Becker, A., D. Geiger, and C. Meek (2000). Perfect tree-like Markovian distributions. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 19–23.

Becker, A., D. Geiger, and A. Schäffer (1998). Automatic selection of loop breakers for genetic linkage analysis. *Human Heredity 48*, 49–60.

Beeri, C., R. Fagin, D. Maier, and M. Yannakakis (1983). On the desirability of acyclic database schemes. *Journal of the Association for Computing Machinery 30*(3), 479–513.

Beinlich, L., H. Suermondt, R. Chavez, and G. Cooper (1989). The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine*, pp. 247–256. Springer Verlag.

Bell, D. (1982). egret in decision making under uncertainty. *Operations Research 30*, 961–981.

Bellman, R. E. (1957). *Dynamic Programming*. Princeton, New Jersey: Princeton University Press.

Ben-Tal, A. and A. Charnes (1979). A dual optimization framework for some problems of information theory and statistics. *Problems of Control and Information Theory 8*, 387–401.

Bentham, J. (1789). An introduction to the principles of morals and legislation.

Berger, A., S. Della-Pietra, and V. Della-Pietra (1996). A maximum entropy approach to natural language processing. *Computational Linguistics 16*(2).

Bernardo, J. and A. Smith (1994). *Bayesian Theory*. New York: John Wiley and Sons.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis (exposition of a new theory on the measurement of risk). English Translation by L. Sommer, *Econometrica*, 22:23–36, 1954.

Berrou, C., A. Glavieux, and P. Thitimajshima (1993). Near Shannon limit error-correcting coding: Turbo codes. In *Proc. International Conference on Communications*, pp. 1064–1070.

Bertelé, U. and F. Brioschi (1972). *Nonserial Dynamic Programming*. New York: Academic Press.

Bertsekas, D. (1999). *Nonlinear Programming* (2nd ed.). Athena Scientific.

Bertsekas, D. P. and J. N. Tsitsiklis (1996). *Neuro-Dynamic Programming*. Athena Scientific.

Besag, J. (1977a). Efficiency of pseudo-likelihood estimation for simple Gaussian fields. *Biometrika 64*(3), 616–618.

Besag, J. (1977b). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B 36*, 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B 48*, 259–302.

Bethe, H. A. (1935). Statistical theory of superlattices. *in Proceedings of the Royal Society of London A*, 552.

Bidyuk, B. and R. Dechter (2007). Cutset sampling for bayesian networks. *Journal of Artificial Intelligence Research 28*, 1–48.

Bilmes, J. and C. Bartels (2003). On triangulating dynamic graphical models. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Bilmes, J. and C. Bartels (2005, September). Graphical model architectures for speech recognition. *IEEE Signal Processing Magazine 22*(5), 89–100.

Binder, J., D. Koller, S. Russell, and K. Kanazawa (1997). Adaptive probabilistic networks with hidden variables. *Machine Learning 29*, 213–244.

Binder, J., K. Murphy, and S. Russell (1997). Space-efficient inference in dynamic probabilistic networks. In *Proc. 15th International Joint Conference on Artificial Intelligence (IJCAI)*.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics

(M. Jordan, J. Kleinberg, and B. Schökopf, editors). New York: Springer-Verlag.

Bishop, C., N. Lawrence, T. Jaakkola, and M. Jordan (1997). Approximating posterior distributions in belief networks using mixtures. In *Proc. 11th Conference on Neural Information Processing Systems (NIPS)*.

Blalock, Jr., H. (1971). *Causal Models in the Social Sciences*. Chicago, Illinois: Aldine-Atheson.

Blum, B., C. Shelton, and D. Koller (2006). A continuation method for nash equilibria in structured games. *Journal of Artificial Intelligence Resarch 25*, 457–502.

Bodlaender, H., A. Koster, F. van den Eijkhof, and L. van der Gaag (2001). Pre-processing for triangulation of probabilistic networks. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 32–39.

Boros, E. and P. Hammer (2002). Pseudo-Boolean optimization. *Discrete Applied Mathematics 123*(1-3).

Bouckaert, R. (1993). Probabilistic network construction using the minimum description length principle. In *Proc. European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 41–48.

Boutilier, C. (2002). A POMDP formulation of preference elicitation problems. In *Proc. 18th Conference on Artificial Intelligence (AAAI)*, pp. 239–46.

Boutilier, C., F. Bacchus, and R. Brafman (2001). UCP-Networks: A directed graphical representation of conditional utilities. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 56–64.

Boutilier, C., T. Dean, and S. Hanks (1999). Decision theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research 11*, 1 – 94.

Boutilier, C., R. Dearden, and M. Goldszmidt (1989). Exploiting structure in policy construction. In *Proc. 14th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1104–1111.

Boutilier, C., R. Dearden, and M. Goldszmidt (2000). Stochastic dynamic programming with factored representations. *Artificial Intelligence 121*(1), 49–107.

Boutilier, C., N. Friedman, M. Goldszmidt, and D. Koller (1996). Context-specific independence in Bayesian networks. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 115–123.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Boyen, X., N. Friedman, and D. Koller (1999). Discovering the hidden structure of complex dynamic systems. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 91–100.

Boyen, X. and D. Koller (1998a). Approximate learning of dynamic models. In *Proc. 12th Conference on Neural Information Processing Systems (NIPS)*.

Boyen, X. and D. Koller (1998b). Tractable inference for complex stochastic processes. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 33–42.

Boyen, X. and D. Koller (1999). Exploiting the architecture of dynamic systems. In *Proc. 15th Conference on Artificial Intelligence (AAAI)*.

Boykov, Y., O. Veksler, and R. Zabih (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*(11), 1222–1239.

Braziunas, D. and C. Boutilier (2005). Local utility elicitation in GAI models. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 42–49.

Breese, J. and D. Heckerman (1996). Decision-theoretic troubleshooting: A framework for repair and experiment. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp.

124–132.

Breese, J., D. Heckerman, and C. Kadie (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 43–52.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees*. Monterey,CA: Wadsworth & Brooks.

Buchanan, B. and E. Shortliffe (Eds.) (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley.

Bui, H., S. Venkatesh, and G. West (2001). Tracking and surveillance in wide-area spatial environments using the Abstract Hidden Markov Model. *International Journal of Pattern Recognition and Artificial Intelligence*.

Buntine, W. (1991). Theory refinement on Bayesian networks. In *Proc. 7th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 52–60.

Buntine, W. (1993). Learning classification trees. In D. J. Hand (Ed.), *Artificial Intelligence Frontiers in Statistics*, Number III in AI and Statistics. Chapman & Hall.

Buntine, W. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research 2*, 159–225.

Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on Knowledge and Data Engineering 8*, 195–210.

Caffo, B., W. Jank, and G. Jones (2005). Ascent-based Monte Carlo Expectation-Maximization. *Journal of the Royal Statistical Society, Series B*.

Cannings, C., E. A. Thompson, and H. H. Skolnick (1976). The recursive derivation of likelihoods on complex pedigrees. *Advances in Applied Probability 8*(4), 622–625.

Cannings, C., E. A. Thompson, and M. H. Skolnick (1978). Probability functions on complex pedigrees. *Advances in Applied Probability 10*(1), 26–61.

Cano, J., L.D., Hernández, and S. Moral (2006). Importance sampling algorithms for the propagation of probabilities in belief networks. *International Journal of Approximate Reasoning 15*(1), 77–92.

Carreira-Perpignan, M. and G. Hinton (2005). On contrastive divergence learning. In *Proc. 11thWorkshop on Artificial Intelligence and Statistics*.

Casella, G. and R. Berger (1990). *Statistical Inference*. Wadsworth.

Castillo, E., J. Gutiérrez, and A. Hadi (1997a). *Expert Systems and Probabilistic Network Models*. New York: Springer-Verlag.

Castillo, E., J. Gutiérrez, and A. Hadi (1997b). Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics 27*, 412–23.

Chajewska, U. (2002). *Acting Rationally with Incomplete Utility Information*. Ph.D. thesis, Stanford University.

Chajewska, U. and D. Koller (2000). Utilities as random variables: Density estimation and structure discovery. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 63–71.

Chajewska, U., D. Koller, and R. Parr (2000). Making rational decisions using adaptive utility elicitation. In *Proc. 16th Conference on Artificial Intelligence (AAAI)*, pp. 363–369.

Chan, H. and A. Darwiche (2002). When do numbers really matter? *Journal of Artificial Intelligence Research 17*, 265–287.

Chávez, T. and M. Henrion (1994). Efficient estimation of the value of information in Monte Carlo

models. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 119–127.

Cheeseman, P., J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman (1988). Autoclass: a Bayesian classification system. In *Proc. 5th International Conference on Machine Learning (ICML)*.

Cheeseman, P., M. Self, J. Kelly, and J. Stutz (1988). Bayesian classification. In *Proc. 4th Conference on Artificial Intelligence (AAAI)*, Volume 2, pp. 607–611.

Cheeseman, P. and J. Stutz (1995). Bayesian classification (AutoClass): Theory and results. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*. AAAI Press.

Chen, L., M. Wainwright, M. Cetin, and A. Willsky (2003). Multitarget-multisensor data association using the tree-reweighted max-product algorithm. In *Proceedings SPIE Aerosense Conference*, Orlando, Florida.

Chen, R. and S. Liu (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society, Series B.*

Cheng, J. and M. Druzdzel (2000). AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. *Journal of Artificial Intelligence Research 13*, 155–188.

Cheng, J., R. Greiner, J. Kelly, D. Bell, and W. Liu (2002). Learning bayesian networks from data: An information-theory based approach. *Artificial Intelligence.*

Chesley, G. (1978). Subjective probability elicitation techniques: A performance comparison. *Journal of Accounting Research 16*(2), 225–241.

Chickering, D. (1996a). Learning Bayesian networks is NP-Complete. In D. Fisher and H. Lenz (Eds.), *Learning from Data: Artificial Intelligence and Statistics V*, pp. 121–130. Springer-Verlag.

Chickering, D. (2002a, February). Learning equivalence classes of Bayesian-network structures. *Journal of Machine Learning Research 2*, 445–498.

Chickering, D., D. Geiger, and D. Heckerman (1995, January). Learning Bayesian networks: Search methods and experimental results. In *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, pp. 112–128.

Chickering, D., C. Meek, and D. Heckerman (2003). Large-sample learning of Bayesian networks is hard. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 124–133.

Chickering, D. and J. Pearl (1997). A clinician's tool for analyzing non-compliance. *Computing Science and Statistics 29*, 424–31.

Chickering, D. M. (1995). A transformational characterization of equivalent Bayesian network structures. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 87–98.

Chickering, D. M. (1996b). Learning equivalence classes of Bayesian network structures. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 150–157.

Chickering, D. M. (2002b, November). Optimal structure identification with greedy search. *Journal of Machine Learning Research 3*, 507–554.

Chickering, D. M. and D. Heckerman (1997). Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning 29*, 181–212.

Chickering, D. M., D. Heckerman, and C. Meek (1997). A Bayesian approach to learning Bayesian networks with local structure. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 80–89.

Chow, C. K. and C. N. Liu (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory 14*, 462–467.

Collins, M. (2002). Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proc. Conference on Empirical Methods in Natural*

Language Processing (EMNLP).

Cooper, G. (1990). Probabilistic inference using belief networks is NP-hard. *Artificial Intelligence 42*, 393–405.

Cooper, G. and E. Herskovits (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning 9*, 309–347.

Cooper, G. and C. Yoo (1999). Causal discovery from a mixture of experimental and observational data. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 116–125.

Cooper, G. F. (1988). A method for using belief networks as influence diagrams. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence (UAI)*, pp. 55–63.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2001). *Introduction to Algorithms*. Cambridge, Massachusetts: MIT Press. 2nd Edition.

Covaliu, Z. and R. Oliver (1995). Representation and solution of decision problems using sequential decision diagrams. *Management Science 41*(12), 1860–81.

Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory*. John Wiley & Sons.

Cowell, R. (2005). Local propagation in conditional gaussian Bayesian networks. *Journal of Machine Learning Research 6*, 1517–1550.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag.

Cox, R. (2001). *Algebra of Probable Inference*. The Johns Hopkins University Press.

Cozman, F. (2000). Credal networks. *Artificial Intelligence 120*, 199–233.

Csiszàr, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability 3*(1), 146–158.

Culotta, A., M. Wick, R. Hall, and A. McCallum (2007). First-order probabilistic models for coreference resolution. In *Proc. Conference of the North American Association for Computational Linguistics*.

D. Rusakov, D. G. (2005). Asymptotic model selection for naive Bayesian networks. *Journal of Machine Learning Research 6*, 1–35.

Dagum, P. and M. Luby (1993). Appoximating probabilistic inference in Bayesian belief networks in NP-hard. *Artificial Intelligence 60*(1), 141–153.

Dagum, P. and M. Luby (1997). An optimal approximation algorithm for Baysian inference. *Artificial Intelligence 93*(1–2), 1–27.

Daneshkhah, A. (2004). Psychological aspects influencing elicitation of subjective probability. Technical report, University of Sheffield.

Darroch, J. and D. Ratcliff (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics 43*, 1470–1480.

Darwiche, A. (1993). Argument calculus and networks. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 420–27.

Darwiche, A. (2001a). Constant space reasoning in dynamic Bayesian networks. *International Journal of Approximate Reasoning 26*, 161–178.

Darwiche, A. (2001b). Recursive conditioning. *Artificial Intelligence 125*(1–2), 5–41.

Darwiche, A. (2003). A differential approach to inference in Bayesian networks. *Journal of the ACM 50*(3), 280–305.

Darwiche, A. and M. Goldszmidt (1994). On the relation between Kappa calculus and probabilistic reasoning. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Dasgupta, S. (1997). The sample complexity of learning fixed-structure Bayesian networks. *Ma-*

chine Learning 29, 165–180.

Dasgupta, S. (1999). Learning polytrees. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 134–141.

Dawid, A. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society, Series B 41*, 1–31.

Dawid, A. (1980). Conditional independence for statistical operations. *Annals of Statistics 8*, 598–617.

Dawid, A. (1984). Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Series A 147*(2), 278–292.

Dawid, A. (1992). Applications of a general propagation algorithm for probabilistic expert system. *Statistics and Computing 2*, 25–36.

Dawid, A. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review 70*, 161–189. Corrections p437.

Dawid, A. (2007, September). Fundamentals of statistical causality. Technical Report 279, RSS/EPSRC Graduate Training Programme, University of Sheffield.

Dawid, A., U. Kjærulff, and S. Lauritzen (1995). Hybrid propagation in junction trees. In *Advances in Intelligent Computing*, Volume 945. Springer-Verlag.

de Bombal, F., D. Leaper, J. Staniland, A. McCann, and J. Harrocks (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal 2*, 9–13.

de Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. *Annals Institute H. Poincaré 7*, 1–68. Translated by H. Kyburg in Kyburg et al. (1980).

de Freitas, N., P. Højen-Sørensen, M. Jordan, and S. Russell (2001). Variational MCMC. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 120–127.

Dean, T. and K. Kanazawa (1989). A model for reasoning about persistence and causation. *Computational Intelligence 5*(3), 142–150.

Dechter, R. (1997). Mini-Buckets: A general scheme for generating approximations in automated reasoning. In *Proc. 15th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1297–1303.

Dechter, R. (1999). Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence 113*(1–2), 41–85.

Dechter, R. (2003). *Constraint Processing*. Morgan Kaufmann.

Dechter, R., K. Kask, and R. Mateescu (2002). Iterative join-graph propagation. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 128–136.

Dechter, R. and I. Rish (1997). A scheme for approximating probabilistic inference. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*.

DeGroot, M. H. (1989). *Probability and Statistics*. Reading, MA: Addison Wesley.

Della Pietra, S., V. Della Pietra, and J. Lafferty (1997). Inducing features of random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence 19*(4), 380–393.

Dellaert, F., S. Seitz, C. Thorpe, and S. Thrun (2003). EM, MCMC, and chain flipping for structure from motion with unknown correspondence. *Machine Learning 50*(1–2), 45–71.

Deming, W. and F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics 11*, 427–444.

Dempster, A., N. M. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*(1), 1–22.

Deng, K. and A. Moore (1989). Multiresolution instance-based learning. In *Proc. 14th International*

Joint Conference on Artificial Intelligence (IJCAI), pp. 1233–1239.

Deshpande, A., M. Garofalakis, and M. Jordan (2001). Efficient stepwise selection in decomposable models. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 128–135.

Diez, F. (1993). Parameter adjustment in Bayes networks: The generalized noisy OR-gate. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 99–105.

Dittmer, S. L. and F. V. Jensen (1997). Myopic value of information in influence diagrams. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 142–149.

Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering. Technical Report CUED/FINFENG/TR 310, Department of Engineering, Cambridge University.

Doucet, A., N. de Freitas, and N. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag.

Doucet, A., N. de Freitas, K. Murphy, and S. Russell (2000). Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Doucet, A., S. Godsill, and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing 10*(3), 197–208.

Drummond, M., B. O'Brien, G. Stoddart, and G. Torrance (1997). *Methods for the Economic Evaluation of Health Care Programmes, 2nd Edition.* Oxford, UK: Oxford University Press.

Druzdzel, M. (1993). *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense.* Ph.D. thesis, Carnegie Mellon University.

Dubois, D. and H. Prade (1990). Inference i possibilistic hypergraphs. In *Proc. of the 6th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems.*

Duchi, J., D. Tarlow, G. Elidan, and D. Koller (2006). Using combinatorial optimization within max-product belief propagation. In *Proc. 20th Conference on Neural Information Processing Systems (NIPS)*.

Duda, R., J. Gaschnig, and P. Hart (1979). Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie (Ed.), *Expert Systems in the Microelectronic Age*, pp. 153–167. Edinburgh, Scotland: Edinburgh University Press.

Duda, R. and P. Hart (1973). *Pattern Classification and Scene Analysis.* New York: John Wiley & Sons.

Duda, R., P. Hart, and D. Stork (2000). *Pattern Classification, Second Edition.* Wiley.

Dudík, M., S. Phillips, and R. Schapire (2004). Performance guarantees for regularized maximum entropy density estimation. In *Proc. Conference on Computational Learning Theory (COLT)*.

Durbin, R., S. Eddy, A. Krogh, and G. Mitchison (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge: Cambridge University Press.

Dykstra, R. and J. Lemke (1988). Duality of I projections and maximum likelihood estimation for log-linear models under cone constraints. *Journal of the American Statistical Association 83*(402), 546–554.

El-Hay, T. and N. Friedman (2001). Incorporating expressive graphical models in variational approximations: Chain-graphs and hidden variables. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 136–143.

Elfadel, I. (1995). Convex potentials and their conjugates in analog mean-field optimization. *Neural Computation 7*, 1079–1104.

Elidan, G. and N. Friedman (2005). Learning hidden variable networks: The information bottleneck approach. *Journal of Machine Learning Research 6*, 81–127.

Elidan, G., I. McGraw, and D. Koller (2006). Residual belief propagation: Informed scheduling for asynchronous message passing. In *Proc. 22nd Conf. on Uncertainty in Artificial Intelligence*.

Elidan, G., N. Lotner, N. Friedman, and D. Koller (2000). Discovering hidden variables: A structure-based approach. In *Proc. 14th Conf. on Neural Information Processing Systems (NIPS)*.

Elidan, G., I. Nachman, and N. Friedman (2007). "Ideal Parent" structure learning for continuous variable networks. *Journal of Machine Learning Research 8*, 1799–1833.

Elidan, G., M. Ninio, N. Friedman, and D. Schuurmans (2002). Data perturbation for escaping local maxima in learning. In *Proc. 18th National Conference on Artificial Intelligence (AAAI)*.

Ellis, B. and W. Wong (2008). Learning causal Bayesian network structures from experimental data. *Journal of the American Statistical Association 103*, 778–789.

Elston, R. C. and J. Stewart (1971). A general model for the analysis of pedigree data. *Human Heredity 21*, 523–542.

Ezawa, K. (1994). Value of evidence on influence diagrams. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 212–220.

Feller, W. (1970). *An Introduction to Probability Theory and Its Applications* (third ed.), Volume I. New York: John Wiley & Sons.

Felzenszwalb, P. and D. Huttenlocher (2006, October). Efficient belief propagation for early vision. *International Journal of Computer Vision 70*(1).

Fertig, K. and J. Breese (1989). Interval influence diagrams. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Fine, S., Y. Singer, and N. Tishby (1998). The hierarchical Hidden Markov Model: Analysis and applications. *Machine Learning 32*, 41–62.

Fishburn, P. (1967). Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review 8*, 335–42.

Fishburn, P. (1970). *Utility Theory for Decision Making*. New York: Wiley.

Fishelson, M. and D. Geiger (2003). Optimizing exact genetic linkage computations. In *Proc. International Conf. on Research in Computational Molecular Biology (RECOMB)*, pp. 114–121.

Fishman, G. (1976, July). Sampling from the gamma distribution on a computer. *Communications of the ACM 19*(7), 407–409.

Fishman, G. (1996). *Monte Carlo — Concept, Algorithms, and Applications*. Series in Operations Research. Springer.

Fox, D., W. Burgard, and S. Thrun (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research 11*, 391–427.

Freund, Y. and R. Schapire (1998). Large margin classification using the perceptron algorithm. In *Proc. Conference on Computational Learning Theory (COLT)*.

Frey, B. (2003). Extending factor graphs so as to unify directed and undirected graphical models. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 257–264.

Frey, B. and A. Kannan (2000). Accumulator networks: suitors of local probability propagation. In *Proc. 14th Conference on Neural Information Processing Systems (NIPS)*.

Frey, B. and D. MacKay (1997). A revolution: Belief propagation in graphs with cycles. In *Proc. 11th Conference on Neural Information Processing Systems (NIPS)*.

Frey, B. J. (1998). *Graphical Models for Machine Learning and Digital Communication*. Cambridge, Massachusetts: MIT Press.

Friedman, N. (1997). Learning belief networks in the presence of missing values and hidden variables. In *Proc. 14th International Conference on Machine Learning (ICML)*, pp. 125–133.

Friedman, N. (1998). The Bayesian structural em algorithm. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 129–138.

Friedman, N., D. Geiger, and M. Goldszmidt (1997). Bayesian network classifiers. *Machine Learning 29*, 131–163.

Friedman, N., D. Geiger, and N. Lotner (2000). Likelihood computations using value abstraction. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Friedman, N., L. Getoor, D. Koller, and A. Pfeffer (1999). Learning probabilistic relational models. In *Proc. 16th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1300–1307.

Friedman, N. and M. Goldszmidt (1996). Learning Bayesian networks with local structure. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 252–262.

Friedman, N. and M. Goldszmidt (1998). Learning Bayesian networks with local structure. See Jordan (1998), pp. 421–460.

Friedman, N. and D. Koller (2003). Being Bayesian about Bayesian network structure: A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning 50*(1–2), 95–126.

Friedman, N., K. Murphy, and S. Russell (1998). Learning the structure of dynamic probabilistic networks. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Friedman, N. and I. Nachman (2000). Gaussian process networks. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 211–219.

Friedman, N. and Z. Yakhini (1996). On the sample complexity of learning Bayesian networks. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Frogner, C. and A. Pfeffer (2007). Discovering weakly-interacting factors in a complex stochastic process. In *Proc. 21st Conference on Neural Information Processing Systems (NIPS)*.

Frydenberg, J. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics 17*, 790–805.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. MIT Press.

Fung, R. and K. C. Chang (1989). Weighting and integrating evidence for stochastic simulation in Bayesian networks. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, San Mateo, California. Morgan Kaufmann.

Fung, R. and B. del Favero (1994). Backward simulation in Bayesian networks. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 227–234.

Galles, D. and J. Pearl (1995). Testing identifiability of causal models. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 185–95.

Gamerman, D. and H. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, CRC.

Ganapathi, V., D. Vickrey, J. Duchi, and D. Koller (2008). Constrained approximate maximum entropy learning. In *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Garcia, L. D. (2004). Algebraic statistics in model selection. In *Proc. 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 177–18.

Geiger, D. and D. Heckerman. A characterization of the bivariate normal-Wishart distribution. *Probability and Mathematical Statistics 18*, 119–131.

Geiger, D. and D. Heckerman (1994). Learning gaussian networks. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 235–243.

Geiger, D. and D. Heckerman (1996). Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence 82*(1-2), 45–74.

Geiger, D., D. Heckerman, H. King, and C. Meek (2001). Stratified exponential families: Graphical

models and model selection. *Annals of Statistics 29*, 505–529.

Geiger, D., D. Heckerman, and C. Meek (1996). Asymptotic model selection for directed networks with hidden variables. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 283–290.

Geiger, D. and C. Meek (1998). Graphical models and exponential families. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 156–165.

Geiger, D., C. Meek, and Y. Wexler (2006). A variational inference procedure allowing internal structure for overlapping clusters and deterministic constraints. *Journal of Artificial Intelligence Research 27*, 1–23.

Geiger, D. and J. Pearl (1988). On the logic of causal models. In *Proc. 4th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 3–14.

Geiger, D. and J. Pearl (1993). Logical and algorithmic properties of conditional independence and graphical models. *Annals of Statistics 21*(4), 2001–21.

Geiger, D., T. Verma, and J. Pearl (1989). d-separation: From theorems to algorithms. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 139–148.

Geiger, D., T. Verma, and J. Pearl (1990). Identifying independence in Bayesian networks. *Networks 20*, 507–534.

Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Gelman, A. and X.-L. Meng (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science 13*(2), 163–185.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–511.

Geman, S. and D. Geman (1984, November). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence 6*(6), 721–741.

Getoor, L., N. Friedman, D. Koller, A. Pfeffer, and B. Taskar (2007). Probabilistic relational models. See Getoor and Taskar (2007).

Getoor, L., N. Friedman, D. Koller, and B. Taskar (2002). Learning probabilistic models of link structure. *Journal of Machine Learning Research 3*(December), 679–707.

Getoor, L. and B. Taskar (Eds.) (2007). *Introduction to Statistical Relational Learning*. MIT Press.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–1339.

Geyer, C. and E. Thompson (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B*.

Geyer, C. and E. Thompson (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association 90*(431), 909–920.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface Interface Foundation*, pp. 156–163. Fairfax Station.

Ghahramani, Z. (1994). Factorial learning and the em algorithm. In *Proc. 8th Conference on Neural Information Processing Systems (NIPS)*, pp. 617–624.

Ghahramani, Z. and M. Beal (2000). Propagation algorithms for variational Bayesian learning. In

Proc. 14th Conference on Neural Information Processing Systems (NIPS).

Ghahramani, Z. and G. Hinton (1998). Variational learning for switching state-space models. *Neural Computation 12*(4), 963–996.

Ghahramani, Z. and M. Jordan (1993). Supervised learning from incomplete data via an EM approach. In *Proc. 7th Conference on Neural Information Processing Systems (NIPS).*

Ghahramani, Z. and M. Jordan (1997). Factorial hidden Markov models. *Machine Learning 29*, 245–273.

Gibbs, J. (1902). *Elementary Principles of Statistical Mechanics.* New Haven, Connecticut: Yale University Press.

Gidas, B. (1988). Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbsian distributions. In W. Fleming and P.-L. Lions (Eds.), *Stochastic differential systems, stochastic control theory and applications.* Springer, New York.

Gilks, W. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 641–649. Oxford, UK: Clarendon Press.

Gilks, W., N. Best, and K. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Annals of Statistics 44*, 455–472.

Gilks, W., S. Richardson, and D. Spiegelhalter (Eds.) (1996). *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London.

Gilks, W., A. Thomas, and D. Spiegelhalter (1994). A language and program for complex Bayesian modeling. *The Statistician 43*, 169–177.

Gilks, W. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Annals of Statistics 41*, 337–348.

Giudici, P. and P. Green (1999, December). Decomposable graphical Gaussian model determination. *Biometrika 86*(4), 785–801.

Globerson, A. and T. Jaakkola (2007a). Convergent propagation algorithms via oriented trees. In *Proc. 23rd Conference on Uncertainty in Artificial Intelligence (UAI).*

Globerson, A. and T. Jaakkola (2007b). Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *Proc. 21st Conference on Neural Information Processing Systems (NIPS).*

Glover, F. and M. Laguna (1993). Tabu search. In C. Reeves (Ed.), *Modern Heuristic Techniques for Combinatorial Problems*, Oxford, England. Blackwell Scientific Publishing.

Glymour, C. and G. F. Cooper (Eds.) (1999). *Computation, Causation, Discovery.* Cambridge: MIT Press.

Godsill, S., A. Doucet, and M. West (2000). Methodology for Monte Carlo smoothing with application to time-varying autoregressions. In *Proc. International Symposium on Frontiers of Time Series Modelling.*

Golumbic, M. (1980). *Algorithmic Graph Theory and Perfect Graphs.* London: Academic Press.

Good, I. (1950). *Probability and the Weighing of Evidence.* London: Griffin.

Goodman, J. (2004). Exponential priors for maximum entropy models. In *Proc. Conference of the North American Association for Computational Linguistics.*

Goodman, L. (1970). The multivariate analysis of qualitative data: Interaction among multiple classification. *Journal of the American Statistical Association 65*, 226–56.

Gordon, N., D. Salmond, and A. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F 140*(2), 107–113.

Gorry, G. and G. Barnett (1968). Experience with a model of sequential diagnosis. *Computers and Biomedical Research 1*, 490–507.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Green, P. J. (1990). On use of the EM algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society, Series B 52*(3), 443–452.

Greig, D., B. Porteous, and A. Seheult (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B 51*(2), 271–279.

Greiner, R. and W. Zhou (2002). Structural extension to logistic regression: Discriminant parameter learning of belief net classifiers. In *Proc. 18th Conference on Artificial Intelligence (AAAI)*.

Guestrin, C. E., D. Koller, R. Parr, and S. Venkataraman (2003). Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research 19*, 399–468.

Guyon, X. and H. R. Künsch (1992). Asymptotic comparison of estimators in the Ising model. In *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis, Lecture Notes in Statistics*, Volume 74, pp. 177–198. Springer, Berlin.

Ha, V. and P. Haddawy (1997). Problem-focused incremental elicitation of multi-attribute utility models. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 215–222.

Ha, V. and P. Haddawy (1999). A hybrid approach to reasoning with partially elicited preference models. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 263–270.

Haberman, S. (1974). *The General Log-Linear Model*. Ph.D. thesis, Department of Statistics, University of Chicago.

Halpern, J. Y. (2003). *Reasoning about Uncertainty*. MIT Press.

Hammer, P. (1965). Some network flow problems solved with pseudo-Boolean programming. *Operations Research 13*, 388–399.

Hammersley, J. and P. Clifford (1971). Markov fields on finite graphs and lattices. Unpublished manuscript.

Handschin, J. and D. Mayne (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control 9*(5), 547–559.

Hartemink, A., D. Gifford, T. Jaakkola, and R. Young (2002, March/April). Bayesian methods for elucidating genetic regulatory networks. *IEEE Intelligent Systems 17*, 37–43. special issue on Intelligent Systems in Biology.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Springer Series in Statistics.

Hazan, T. and A. Shashua (2008). Convergent message-passing algorithms for inference over general graphs with convex free energies. In *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Heckerman, D. (1990). *Probabilistic Similarity Networks*. MIT Press.

Heckerman, D. (1993). Causal independence for knowledge acquisition and inference. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 122–127.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. See Jordan (1998).

Heckerman, D. and J. Breese (1996). Causal independence for probability assessment and inference using Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics 26*, 826–831.

Heckerman, D., J. Breese, and K. Rommelse (1995, March). Decision-theoretic troubleshooting.

Communications of the ACM 38(3), 49–57.

Heckerman, D., D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie (2000). Dependency networks for inference, collaborative filtering, and data visualization. *jmlr 1*, 49–75.

Heckerman, D. and D. Geiger (1995). Learning Bayesian networks: a unification for discrete and Gaussian domains. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 274–284.

Heckerman, D., D. Geiger, and D. M. Chickering (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning 20*, 197–243.

Heckerman, D., E. Horvitz, and B. Nathwani (1992). Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine 31*, 90–105.

Heckerman, D. and H. Jimison (1989). A Bayesian perspective on confidence. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 149–160.

Heckerman, D., A. Mamdani, and M. Wellman (1995). Real-world applications of Bayesian networks. *Communications of the ACM 38*.

Heckerman, D. and C. Meek (1997). Embedded Bayesian network classifiers. Technical Report MSR-TR-97-06, Microsoft Research, Redmond, WA.

Heckerman, D., C. Meek, and G. Cooper (1999). A Bayesian approach to causal discovery. See Glymour and Cooper (1999), pp. 141–166.

Heckerman, D., C. Meek, and D. Koller (2007). Probabilistic entity-relationship models, PRMs, and plate models. See Getoor and Taskar (2007).

Heckerman, D. and B. Nathwani (1992a). An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research 25(1)*, 56–74.

Heckerman, D. and B. Nathwani (1992b). Toward normative expert systems. II. Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine 31*, 106–16.

Heckerman, D. and R. Shachter (1994). A decision-based view of causality. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 302–310. Morgan Kaufmann.

Henrion, M. (1986). Propagation of uncertainty in Bayesian networks by probabilistic logic sampling. In *Proc. 2nd Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 149–163.

Henrion, M. (1991). Search-based algorithms to bound diagnostic probabilities in very large belief networks. In *Proc. 7th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 142–150.

Hernández, L. and S. Moral (1997). Mixing exact and importance sampling propagation algorithms in dependence graphs. *International Journal of Intelligent Systems 12*, 553–576.

Heskes, T. (2002). Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In *Proc. 16th Conference on Neural Information Processing Systems (NIPS)*, pp. 359–366.

Heskes, T. (2004). On the uniqueness of loopy belief propagation fixed points. *Neural Computation 16*, 2379–2413.

Heskes, T. (2006). Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies. *Journal of Machine Learning Research 26*, 153–190.

Heskes, T., K. Albers, and B. Kappen (2003). Approximate inference and constrained optimization. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 313–320.

Heskes, T., M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter (2005). Approximate inference techniques with expectation constraints. *Journal of Statistical Mechanics: Theory and Experiment*.

Heskes, T. and O. Zoeter (2002). Expectation propagation for approximate inference in dynamic

Bayesian networks. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Heskes, T. and O. Zoeter (2003). Generalized belief propagation for approximate inference in hybrid Bayesian networks. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*.

Heskes, T., O. Zoeter, and W. Wiegerinck (2003). Approximate expectation maximization. In *Proc. 17th Conference on Neural Information Processing Systems (NIPS)*, pp. 353–360.

Higdon, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association 93*, 585–595.

Hinton, G. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation 14*, 1771–1800.

Hinton, G., S. Osindero, and Y. Teh (2006). A fast learning algorithm for deep belief nets. *Neural Computation 18*, 1527–1554.

Hinton, G. and R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science 313*, 504–507.

Hinton, G. and T. Sejnowski (1983). Optimal perceptual inference. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 448–453.

Hinton, G. E., P. Dayan, B. Frey, and R. M. Neal (1995). The wake-sleep algorithm for unsupervised neural networks. *Science 268*, 1158–1161.

Höffgen, K. (1993). Learning and robust learning of product distributions. In *Proc. Conference on Computational Learning Theory (COLT)*, pp. 77–83.

Hofmann, R. and V. Tresp (1995). Discovering structure in continuous variables using bayesian networks. In *Proc. 9th Conference on Neural Information Processing Systems (NIPS)*.

Horn, G. and R. McEliece (1997). Belief propagation in loopy bayesian networks: experimental results. In *Proceedings if IEEE International Symposium on Information Theory*, pp. 232.

Horvitz, E. and M. Barry (1995). Display of information for time-critical decision making. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 296–305.

Horvitz, E., J. Breese, and M. Henrion (1988). Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning 2*, 247–302. Special Issue on Uncertainty in Artificial Intelligence.

Horvitz, E., H. Suermondt, and G. Cooper (1989). Bounded conditioning: Flexible inference for decisions under scarce resources. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 182–193.

Howard, R. (1970). Decision analysis: Perspectives on inference, decision, and experimentation. *Proceedings of the IEEE 58*, 632–643.

Howard, R. (1977). Risk preference. In R. Howard and J. Matheson (Eds.), *Readings in Decision Analysis*, pp. 429–465. Menlo Park, California: Decision Analysis Group, SRI International.

Howard, R. and J. Matheson (1984a). Influence diagrams. See Howard and Matheson (1984b), pp. 721–762.

Howard, R. and J. Matheson (Eds.) (1984b). *The Principle and Applications of Decision Analysis*. Menlo Park, CA, USA: Strategic Decisions Group.

Howard, R. A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics SSC-2*, 22–26.

Howard, R. A. (1989). Microrisks for medical decision analysis. *International Journal of Technology Assessment in Health Care 5*, 357–370.

Huang, C. and A. Darwiche (1996). Inference in belief networks: A procedural guide. *International*

Journal of Approximate Reasoning 15(3), 225–263.

Huang, F. and Y. Ogata (2002). Generalized pseudo-likelihood estimates for Markov random fields on lattice. *Annals of the Institute of Statistical Mathematics 54*, 1–18.

Ihler, A. (2007). Accuracy bounds for belief propagation. In *Proc. 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

Ihler, A. T., J. W. Fisher, and A. S. Willsky (2003). Message errors in belief propagation. In *Proc. 17th Conference on Neural Information Processing Systems (NIPS)*.

Ihler, A. T., J. W. Fisher, and A. S. Willsky (2005). Loopy belief propagation: Convergence and effects of message errors. *Journal of Machine Learning Research 6*, 905–936.

Imoto, S., S. Kim, T. Goto, S. Aburatani, K. Tashiro, S. Kuhara, and S. Miyano (2003). Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of Bioinformatics and Computational Biology 1*, 231–252.

Indyk, P. (2004). Nearest neighbors in high-dimensional spaces. In J. Goodman and J. O'Rourke (Eds.), *Handbook of Discrete and Computational Geometry* (2nd ed.). CRC Press.

Isard, M. (2003). PAMPAS: Real-valued graphical models for computer vision. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 613–620.

Isard, M. and A. Blake (1998a). Condensation — conditional density propagation for visual tracking. *International Journal of Computer Vision 29*(1), 5–28.

Isard, M. and A. Blake (1998b). A smoothing filter for condensation. In *Proc. European Conference on Computer Vision (ECCV)*, Volume 1, pp. 767–781.

Isham, V. (1981). An introduction to spatial point processes and Markov random fields. *International Statistical Review 49*, 21–43.

Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence 25*(10), 1333–1336.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Z. Phys. 31*, 253–258.

Jaakkola, T. (2001). Tutorial on variational approximation methods. In M. Opper and D. Saad (Eds.), *Advanced mean field methods*, pp. 129–160. Cambridge, Massachusetts: MIT Press.

Jaakkola, T. and M. Jordan (1996a). Computing upper and lower bounds on likelihoods in intractable networks. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 340–348.

Jaakkola, T. and M. Jordan (1996b). Recursive algorithms for approximating probabilities in graphical models. In *Proc. 10th Conference on Neural Information Processing Systems (NIPS)*, pp. 487–93.

Jaakkola, T. and M. Jordan (1997). A variational approach to bayesian logistic regression models and their extensions. In *Proc. 6thWorkshop on Artificial Intelligence and Statistics*.

Jaakkola, T. and M. Jordan (1998). Improving the mean field approximation via the use of mixture models. See Jordan (1998).

Jaakkola, T. and M. Jordan (1999). Variational probabilistic inference and the QMR-DT network. *Journal of Artificial Intelligence Research 10*, 291–322.

Jarzynski, C. (1997, Apr). Nonequilibrium equality for free energy differences. *Physical Review Letters 78*(14), 2690–2693.

Jaynes, E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jensen, F., F. V. Jensen, and S. L. Dittmer (1994). From influence diagrams to junction trees. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 367–73.

Jensen, F. and M. Vomlelová (2003). Unconstrained influence diagrams. In *Proc. 19th Conference*

on Uncertainty in Artificial Intelligence (UAI), pp. 234–41.

Jensen, F. V. (1995). Cautious propagation in Bayesian networks. In Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 323–328.

Jensen, F. V. (1996). An introduction to Bayesian Networks. London: University College London Press.

Jensen, F. V., K. G. Olesen, and S. K. Andersen (1990, August). An algebra of Bayesian belief universes for knowledge-based systems. Networks 20(5), 637–659.

Jerrum, M. and A. Sinclair (1997). The Markov chain Monte Carlo method. In D. Hochbaum (Ed.), Approximation Algorithms for NP-hard Problems. Boston: PWS Publishing.

Ji, C. and L. Seymour (1996). A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood. Annals of Applied Probability.

Jimison, H., L. Fagan, R. Shachter, and E. Shortliffe (1992). Patient-specific explanation in models of chronic disease. AI in Medicine 4, 191–205.

Jordan, M., Z. Ghahramani, T. Jaakkola, and L. K. Saul (1998). An introduction to variational approximations methods for graphical models. See Jordan (1998).

Jordan, M. I. (Ed.) (1998). Learning in Graphics Models. Cambridge, MA: The MIT Press.

Julier, S. (2002). The scaled unscented transformation. In Proceedings of the American Control Conference, Volume 6, pp. 4555–4559.

Julier, S. and J. Uhlmann (1997). A new extension of the Kalman filter to nonlinear systems. In Proc. of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls.

Kahneman, D., P. Slovic, and A. Tversky (Eds.) (1982). Judgment under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Kalman, R. and R. Bucy (1961). New results in linear filtering and prediction theory. Trans. ASME, Series D, Journal of Basic Engineering.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D), 35–45.

Kanazawa, K., D. Koller, and S. Russell (1995). Stochastic simulation algorithms for dynamic probabilistic networks. In Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 346–351.

Kass, R. and A. Raftery (1995). Bayes factors. Journal of the American Statistical Association 90(430), 773–795.

Kearns, M., M. L. Littman, and S. Singh (2001). Graphical models for game theory. In Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 253–260.

Kearns, M. and Y. Mansour (1998). Exact inference of hidden structure from sample data in noisy-or networks. In Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 304–31.

Kearns, M., Y. Mansour, and A. Ng (1997). An information-theoretic analysis of hard and soft assignment methods for clustering. In Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 282–293.

Keeney, R. L. and H. Raiffa (1976). Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, Inc.

Kersting, K. and L. De Raedt (2007). Bayesian logic programming: Theory and tool. See Getoor and Taskar (2007).

Kikuchi, R. (1951). A theory of cooperative phenomena. Physical Review Letters 81, 988–1003.

Kim, C.-J. and C. Nelson (1998). *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press.

Kim, J. and J. Pearl (1983). A computational model for combined causal and diagnostic reasoning in inference systems. In *Proc. 7th International Joint Conference on Artificial Intelligence (IJCAI),* pp. 190–193.

Kirkpatrick, S., C. Gelatt, and M. Vecchi (1983). Optimization by simulated annealing. *Science 220,* 671–680.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics 5*(1), 1–25.

Kjærulff, U. (1990, March). Triangulation of graph — Algorithms giving small total state space. Technical Report R90-09, Aalborg University, Denmark.

Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks. In *Proc. 8th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 121–129.

Kjærulff, U. (1995a). dHugin: A computational system for dynamic time-sliced Bayesian networks. *International Journal of Forecasting 11,* 89–111.

Kjærulff, U. (1995b). HUGS: Combining exact inference and Gibbs sampling in junction trees. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 368–375.

Kjaerulff, U. (1997). Nested junction trees. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 294–301.

Kjærulff, U. and L. van der Gaag (2000). Making sensitivity analysis computationally efficient. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 317–325.

Koivisto, M. and K. Sood (2004). Exact Bayesian structure discovery in Bayesian networks. *Journal of Machine Learning Research 5,* 549–573.

Kok, J., M. Spaan, and N. Vlassis (2003). Multi-robot decision making using coordination graphs. In *Proc. International Conference on Advanced Robotics (ICAR),* pp. 1124–1129.

Kok, J. and N. Vlassis (2005). Using the max-plus algorithm for multiagent decision making in coordination graphs. In *RoboCup-2005: Robot Soccer World Cup IX,* Osaka, Japan.

Koller, D. and R. Fratkina (1998). Using learning for approximation in stochastic processes. In *Proc. 15th International Conference on Machine Learning (ICML),* pp. 287–295.

Koller, D., U. Lerner, and D. Anguelov (1999). A general algorithm for approximate inference and its application to hybrid Bayes nets. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 324–333.

Koller, D. and B. Milch (2001). Multi-agent influence diagrams for representing and solving games. In *Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI),* pp. 1027–1034.

Koller, D. and B. Milch (2003). Multi-agent influence diagrams for representing and solving games. *Games and Economic Behavior 45*(1), 181–221. Full version of paper in IJCAI '03.

Koller, D. and R. Parr (1999). Computing factored value functions for policies in structured MDPs. In *Proc. 16th International Joint Conference on Artificial Intelligence (IJCAI),* pp. 1332–1339.

Koller, D. and A. Pfeffer (1997). Object-oriented Bayesian networks. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI),* pp. 302–313.

Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Kolmogorov, V. and C. Rother (2006). Comparison of energy minimization algorithms for highly connected graphs. In *Proc. European Conference on Computer Vision (ECCV).*

Kolmogorov, V. and M. Wainwright (2005). On the optimality of tree reweighted max-product

message passing. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*.

Kolmogorov, V. and R. Zabih (2004). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*(2).

Komarek, P. and A. Moore (2000). A dynamic adaptation of AD-trees for efficient machine learning on large data sets. In *Proc. 17th International Conference on Machine Learning (ICML)*, pp. 495–502.

Komodakis, N., N. Paragios, and G. Tziritas (2007). MRF optimization via dual decomposition: Message-passing revisited. In *Proc. International Conference on Computer Vision (ICCV)*.

Komodakis, N. and G. Tziritas (2005). A new framework for approximate labeling via graph-cuts. In *Proc. International Conference on Computer Vision (ICCV)*.

Komodakis, N., G. Tziritas, and N. Paragios (2007). Fast, approximately optimal solutions for single and dynamic MRFs. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kong, A. (1991). Efficient methods for computing linkage likelihoods of recessive diseases in inbred pedigrees. *Genetic Epidemiology 8*, 81–103.

Korb, K. and A. Nicholson (2003). *Bayesian Artificial Intelligence*. CRC Press.

Koster, J. (1996). Markov properties of non-recursive causal models. *The Annals of Statistics 24*(5), 2148–77.

Kočka, T., R. Bouckaert, and M. Studený (2001). On characterizing inclusion of Bayesian networks. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 261–68.

Kozlov, A. and D. Koller (1997). Nonuniform dynamic discretization in hybrid networks. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 314–325.

Krause, A. and C. Guestrin (2005a). Near-optimal nonmyopic value of information in graphical models. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*.

Krause, A. and C. Guestrin (2005b). Optimal nonmyopic value of information in graphical models: Efficient algorithms and theoretical limits. In *Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI)*.

Kreps, D. (1988). *Notes on the Theory of Choice*. Boulder, Colorado: Westview Press.

Kschischang, F. and B. Frey (1998). Iterative decoding of compound codes by probability propagation in graphical models. *IEEE Journal on Selected Areas in Communications 16*, 219–230.

Kschischang, F., B. Frey, and H.-A. Loeliger (2001a). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory 47*, 498–519.

Kschischang, F., B. Frey, and H.-A. Loeliger (2001b). Factor graphs and the sum-product algorithm. *IEEE Trans. Information Theory 47*, 498–519.

Kullback, S. (1959). *Information Theory and Statistics*. New York: John Wiley & Sons.

Kumar, M., V. Kolmogorov, and P. Torr (2007). An analysis of convex relaxations for MAP estimation. In *Proc. 21st Conference on Neural Information Processing Systems (NIPS)*.

Kumar, M., P. Torr, and A. Zisserman (2006). Solving Markov random fields using second order cone programming relaxations. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1045–1052.

Kuppermann, M., S. Shiboski, D. Feeny, E. Elkin, and A. Washington (1997, Jan–Mar). Can preference scores for discrete states be used to derive preference scores for an entire path of events? An application to prenatal diagnosis. *Medical Decision Making 17*(1), 42–55.

Kyburg, H., , and H. Smokler (Eds.) (1980). *Studies in Subjective Probability*. New York: Krieger.

La Mura, P. (2000). Game networks. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence*

(UAI), pp. 335–342.

La Mura, P. and Y. Shoham (1999). Expected utility networks. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 366–73.

Lacoste-Julien, S., B. Taskar, D. Klein, and M. Jordan (2006, June). Word alignment via quadratic assignment. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pp. 112–119.

Lafferty, J., A. McCallum, and F. Pereira (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conference on Machine Learning (ICML)*.

Lam, W. and F. Bacchus (1993). Using causal information and local measures to learn Bayesian networks. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 243–250.

Lange, K. and R. C. Elston (1975). Extensions to pedigree analysis. I. Likelihood calculations for simple and complex pedigrees. *Human Heredity 25*, 95–105.

Laskey, K. (1995). Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics 25*(6), 901 – 909.

Lauritzen, S. (1982). *Lectures on contingency tables* (2 ed.). Aalborg: Denmark: University of Aalborg Press.

Lauritzen, S. (1992). Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association 87*(420), 1089–1108.

Lauritzen, S. (1996). *Graphical Models*. New York: Oxford University Press.

Lauritzen, S. and D. Nilsson (2001). Representing and solving decision problems with limited information. *Management Science 47*(9), 1235–51.

Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics and Data Analysis 19*, 191–201.

Lauritzen, S. L. and F. Jensen (2001). Stable local computation with conditional Gaussian distributions. *Statistics and Computing 11*, 191–203.

Lauritzen, S. L. and D. J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B B 50*(2), 157–224.

Lauritzen, S. L. and N. Wermuth (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics 17*, 31–57.

LeCun, Y., S. Chopra, R. Hadsell, R. Marc'Aurelio, and F.-J. Huang (2007). A tutorial on energy-based learning. In G. Bakir, T. Hofmann, B. Schölkopf, A. Smola, B. Taskar, and S. Vishwanathan (Eds.), *Predicting Structured Data*. MIT Press.

Lee, S.-I., V. Ganapathi, and D. Koller (2006). Efficient structure learning of Markov networks using L1-regularization. In *Proc. 20th Conference on Neural Information Processing Systems (NIPS)*.

Lehmann, E. and J. Romano (2008). *Testing Statistical Hypotheses*. Springer Texts in Statistics.

Leisink, M. A. R. and H. J. Kappen (2003). Bound propagation. *Journal of Artificial Intelligence Research 19*, 139–154.

Lerner, U. (2002). *Hybrid Bayesian Networks for Reasoning about Complex Systems*. Ph.D. thesis, Stanford University.

Lerner, U., B. Moses, M. Scott, S. McIlraith, and D. Koller (2002). Monitoring a complex physical system using a hybrid dynamic Bayes net. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 301–310.

Lerner, U. and R. Parr (2001). Inference in hybrid networks: Theoretical limits and practical algorithms. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 310–318.

Lerner, U., R. Parr, D. Koller, and G. Biswas (2000). Bayesian fault detection and diagnosis in dynamic systems. In *Proc. 16th Conference on Artificial Intelligence (AAAI)*, pp. 531–537.

Lerner, U., E. Segal, and D. Koller (2001). Exact inference in networks with discrete children of continuous parents. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 319–328.

Li, S. (2001). *Markov Random Field Modeling in Image Analysis*. Springer.

Liang, P. and M. Jordan (2008). An asymptotic analysis of generative, discriminative, and pseudolikelihood estimators. In *Proc. 25th International Conference on Machine Learning (ICML)*.

Little, R. J. A. (1976). Inference about means for incomplete multivariate data. *Biometrika 63*, 593–604.

Little, R. J. A. and D. B. Rubin (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Liu, D. and J. Nocedal (1989). On the limited memory method for large scale optimization. *Mathematical Programming 45*(3), 503–528.

Liu, J., W. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and sampling schemes. *Biometrika 81*, 27–40.

Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal 92*, 805–824.

MacEachern, S. and L. Berliner (1994, August). Subsampling the Gibbs sampler. *The American Statistician 48*(3), 188–190.

MacKay, D. J. C. (1997). Ensemble learning for hidden markov models. Unpublished manuscripts, http://wol.ra.phy.cam.ac.uk/mackay.

MacKay, D. J. C. and R. M. Neal (1996). Near shannon limit performance of low density parity check codes. *Electronics Letters 32*, 1645–1646.

Madigan, D., S. Andersson, M. Perlman, and C. Volinsky (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic graphs. *Communications in Statistics: Theory and Methods 25*, 2493–2519.

Madigan, D. and E. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association 89*, 1535–1546.

Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International statistical Review 63*, 215–232.

Madsen, A. and D. Nilsson (2001). Solving influence diagrams using HUGIN, Shafer-Shenoy and lazy propagation. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 337–45.

Malioutov, D., J. Johnson, and A. Willsky (2006). Walk-sums and belief propagation in Gaussian graphical models. *Journal of Machine Learning Research 7*, 2031–64.

Maneva, E., E. Mossel, and M. Wainwright (2007, July). A new look at survey propagation and its generalizations. *Journal of the ACM 54*(4), 2–41.

Manning, C. and H. Schuetze (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Marinari, E. and G. Parisi (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters 19*, 451.

Marinescu, R., K. Kask, and R. Dechter (2003). Systematic vs. non-systematic algorithms for solving the MPE task. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Marthi, B., H. Pasula, S. Russell, and Y. Peres (2002). Decayed MCMC filtering. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Martin, J. and K. VanLehn (1995). Discrete factor analysis: Learning hidden variables in Bayesian networks. Technical report, Department of Computer Science, University of Pittsburgh.

McCallum, A. (2003). Efficiently inducing features of conditional random fields. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 403–10.

McCallum, A., C. Pal, G. Druck, and X. Wang (2006). Multi-conditional learning: Generative/discriminative training for clustering and classification. In *Proc. 22nd Conference on Artificial Intelligence (AAAI)*.

McCallum, A. and B. Wellner (2005). Conditional models of identity uncertainty with application to noun coreference. In *Proc. 19th Conference on Neural Information Processing Systems (NIPS)*, pp. 905–912.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. London: Chapman & Hall.

McEliece, R., D. MacKay, and J.-F. Cheng (1998, February). Turbo decoding as an instance of Pearl's "belief propagation" algorithm. *IEEE Journal on Selected Areas in Communications 16*(2).

McEliece, R. J., E. R. Rodemich, and J.-F. Cheng (1995). The turbo decision algorithm. In *Proc. 33rd Allerton Conference on Communication Control and Computing*, pp. 366–379.

McLachlan, G. J. and T. Krishnan (1997). *The EM Algorithm and Extensions*. Wiley Interscience.

Meek, C. (1995a). Causal inference and causal explanation with background knowledge. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 403–418.

Meek, C. (1995b). Strong completeness and faithfulness in Bayesian networks. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 411–418.

Meek, C. (1997). *Graphical Models: Selecting causal and statistical models*. Ph.D. thesis, Carnegie Mellon University.

Meek, C. (2001). Finding a path is harder than finding a tree. *Journal of Artificial Intelligence Research 15*, 383–389.

Meek, C. and D. Heckerman (1997). Structure and parameter learning for causal independence and causal interaction models. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 366–375.

Meila, M. and T. Jaakkola (2000). Tractable Bayesian learning of tree belief networks. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Meila, M. and M. Jordan (2000). Learning with mixtures of trees. *Journal of Machine Learning Research 1*, 1–48.

Meltzer, T., C. Yanover, and Y. Weiss (2005). Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *Proc. International Conference on Computer Vision (ICCV)*, pp. 428–435.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equation of state calculation by fast computing machines. *Journal of Chemical Physics 21*, 1087–1092.

Meyer, J., M. Phillips, P. Cho, I. Kalet, and J. Doctor (2004). Application of influence diagrams to prostate intensity-modulated radiation therapy plan selection. *Physics in Medicine and Biology 49*, 1637–53.

Middleton, B., M. Shwe, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base.

II. Evaluation of diagnostic performance. *Methods of Information in Medicine 30*, 256–67.

Milch, B., B. Marthi, and S. Russell (2004). BLOG: Relational modeling with unknown objects. In *ICML 2004 Workshop on Statistical Relational Learning and Its Connections to Other Fields*.

Milch, B., B. Marthi, S. Russell, D. Sontag, D. Ong, and A. Kolobov (2005). BLOG: Probabilistic models with unknown objects. In *Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1352–1359.

Milch, B., B. Marthi, S. Russell, D. Sontag, D. Ong, and A. Kolobov (2007). BLOG: Probabilistic models with unknown objects. See Getoor and Taskar (2007).

Miller, R., H. Pople, and J. Myers (1982). Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine 307*, 468–76.

Minka, T. (2005). Discriminative models, not discriminative training. Technical Report MSR-TR-2005-144, Microsoft Research.

Minka, T. and J. Lafferty (2002). Expectation propagation for the generative aspect model. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Minka, T. P. (2001a). Algorithms for maximum-likelihood logistic regression. Available from http://www.stat.cmu.edu/~minka/papers/logreg.html.

Minka, T. P. (2001b). Expectation propagation for approximate Bayesian inference. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 362–369.

Møller, J. M., A. Pettitt, K. Berthelsen, and R. Reeves (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalisation constants. *Biometrika 93*(2), 451–458.

Montemerlo, M., S. Thrun, D. Koller, and B. Wegbreit (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proc. 18th Conference on Artificial Intelligence (AAAI)*, pp. 593–598.

Monti, S. and G. F. Cooper (1997). Learning Bayesian belief networks with neural network estimators. In *Proc. 11th Conference on Neural Information Processing Systems (NIPS)*, pp. 579–584.

Mooij, J. M. and H. J. Kappen (2007). Sufficient conditions for convergence of the sum-product algorithm. *IEEE Trans. Information Theory 53*, 4422–4437.

Moore, A. (2000). The anchors hierarchy: Using the triangle inequality to survive high-dimensional data. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 397–405.

Moore, A. and W.-K. Wong (2003). Optimal reinsertion: A new search operator for accelerated and more accurate bayesian network structure learning. In *Proc. 20th International Conference on Machine Learning (ICML)*, pp. 552–559.

Moore, A. W. and M. S. Lee (1997). Cached sufficient statistics for efficient machine learning with large datasets. *Journal of Artificial Intelligence Research 8*, 67–91.

Morgan, M. and M. Henrion (Eds.) (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.

Motwani, R. and P. Raghavan (1995). *Randomized Algorithnms*. Cambridge University Press.

Muramatsu, M. and T. Suzuki (2003). A new second-order cone programming relaxation for max-cut problems. *Journal of Operations Research of Japan 43*, 164–177.

Murphy, K. (1999). Bayesian map learning in dynamic environments. In *Proc. 13th Conference on Neural Information Processing Systems (NIPS)*.

Murphy, K. (2002). Dynamic Bayesian Networks: A tutorial. Technical report, Mas-

sachussetts Institute of Technology. Available from http://www.cs.ubc.ca/~murphyk/Papers/dbnchapter.pdf.

Murphy, K. and M. Paskin (2001). Linear time inference in hierarchical HMMs. In *Proc. 15th Conference on Neural Information Processing Systems (NIPS)*.

Murphy, K. and Y. Weiss (2001). The factored frontier algorithm for approximate inference in DBNs. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Murphy, K. P. (1998). Inference and learning in hybrid Bayesian networks. Technical Report UCB/CSD-98-990, University of California, Berkeley.

Murphy, K. P., Y. Weiss, and M. Jordan (1999). Loopy belief propagation for approximate inference: an empirical study. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 467–475.

Murray, I. and Z. Ghahramani (2004). Bayesian learning in undirected graphical models: Approximate MCMC algorithms. In *Proc. 20th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Murray, I., Z. Ghahramani, and D. MacKay (2006). MCMC for doubly-intractable distributions. In *Proc. 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*.

Myers, J., K. Laskey, and T. Levitt (1999). Learning Bayesian networks from incomplete data with stochastic search algorithms. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 476–485.

Narasimhan, M. and J. Bilmes (2004). PAC-learning bounded tree-width graphical models. In *Proc. 20th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Ndilikilikesha, P. (1994). Potential influence diagrams. *International Journal of Approximate Reasoning 10*, 251–85.

Neal, R. (1996). Sampling from multimodal distributions using tempered transitions. *Statistics and Computing 6*, 353–366.

Neal, R. (2001). Annealed importance sampling. *Statistics and Computing 11*(2), 25–139.

Neal, R. (2003). Slice sampling. *Annals of Statistics 31*(3), 705–767.

Neal, R. M. (1992). Asymmetric parallel Boltzmann machines are belief networks. *Neural Computation 4*(6), 832–834.

Neal, R. M. (1993). Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, University of Toronto.

Neal, R. M. and G. E. Hinton (1998). A new view of the EM algorithm that justifies incremental and other variants. See Jordan (1998).

Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Prentice Hall.

Ng, A. and M. Jordan (2000). Approximate inference algorithms for two-layer Bayesian networks. In *Proc. 14th Conference on Neural Information Processing Systems (NIPS)*.

Ng, A. and M. Jordan (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Proc. 16th Conference on Neural Information Processing Systems (NIPS)*.

Ng, B., L. Peshkin, and A. Pfeffer (2002). Factored particles for scalable monitoring. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 370–377.

Ngo, L. and P. Haddawy (1996). Answering queries from context-sensitive probabilistic knowledge bases. *Theoretical Computer Science*.

Nielsen, J., T. Kočka, and J. M. Peña (2003). On local optima in learning Bayesian networks. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 435–442.

Nielsen, T. and F. Jensen (1999). Welldefined decision scenarios. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 502–11.

Nielsen, T. and F. Jensen (2000). Representing and solving asymmetric Bayesian decision problems. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 416–25.

Nielsen, T., P.-H. Wuillemin, F. Jensen, and U. Kjærulff (2000). Using robdds for inference in Bayesian networks with troubleshooting as an example. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 426–35.

Nilsson, D. (1998). An efficient algorithm for finding the M most probable configurations in probabilistic expert systems. *Statistics and Computing 8*(2), 159–173.

Nilsson, D. and S. Lauritzen (2000). Evaluating influence diagrams with LIMIDs. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 436–445.

Nodelman, U., C. R. Shelton, and D. Koller (2002). Continuous time Bayesian networks. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 378–387.

Nodelman, U., C. R. Shelton, and D. Koller (2003). Learning continuous time Bayesian networks. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Norman, J., Y. Shahar, M. Kuppermann, and B. Gold (1998). Decision-theoretic analysis of prenatal testing strategies. Technical Report SMI-98-0711, Stanford University, Section on Medical Informatics.

Normand, S.-L. and D. Tritchler (1992). Parameter updating in a Bayes network. *Journal of the American Statistical Association 87*, 1109–1115.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.

Nummelin, E. (2002). Mc's for mcmc'ists. *International Statistical Review 70*(2), 215–240.

Olesen, K. G., U. Kjærulff, F. Jensen, B. Falck, S. Andreassen, and S. Andersen (1989). A Munin network for the median nerve — A case study on loops. *Applied Artificial Intelligence 3*, 384–403.

Oliver, R. M. and J. Q. Smith (Eds.) (1990). *Influence Diagrams, Belief Nets and Decision Analysis*. New York: John Wiley & Sons.

Olmsted, S. (1983). *On Representing and Solving Influence Diagrams*. Ph.D. thesis, Stanford University.

Opper, M. and O. Winther (2005). Expectation consistent free energies for approximate inference. In *Proc. 19th Conference on Neural Information Processing Systems (NIPS)*.

Ortiz, L. and L. Kaelbling (1999). Accelerating em: An empirical study. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 512–521.

Ortiz, L. E. and L. P. Kaelbling (2000). Adaptive importance sampling for estimation in structured domains. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 446–454.

Osborne, M. and A. Rubinstein (1994). *A Course in Game Theory*. The MIT Press.

Ostendorf, M., V. Digalakis, and O. Kimball (1996). From HMMs to segment models: A unified view of stochastic modeling for speech recognition. *IEEE Transactions on Speech and Audio Processing 4*(5), 360–378.

Pakzad, P. and V. Anantharam (2002). Minimal graphical representation of Kikuchi regions. In *Proc. 40th Allerton Conference on Communication Control and Computing*, pp. 1585–1594.

Papadimitriou, C. (1993). *Computational Complexity*. Addison Wesley.

Parisi, G. (1988). *Statistical Field Theory*. Reading, Massachusetts: Addison-Wesley.

Park, J. (2002). MAP complexity results and approximation methods. In *Proc. 18th Conference on*

Uncertainty in Artificial Intelligence (UAI), pp. 388–396.

Park, J. and A. Darwiche (2001). Approximating MAP using local search. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 403–âĂŞ410.

Park, J. and A. Darwiche (2003). Solving MAP exactly using systematic search. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Park, J. and A. Darwiche (2004a). Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research 21*, 101–133.

Park, J. and A. Darwiche (2004b). A differential semantics for jointree algorithms. *Artificial Intelligence 156*, 197–216.

Parter, S. (1961). The user of linear graphs in Gauss elimination. *SIAM Review 3*, 119–130.

Paskin, M. (2003a). Sample propagation. In *Proc. 17th Conference on Neural Information Processing Systems (NIPS)*.

Paskin, M. (2003b). Thin junction tree filters for simultaneous localization and mapping. In *Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1157–1164.

Pasula, H., B. Marthi, B. Milch, S. Russell, and I. Shpitser (2002). Identity uncertainty and citation matching. In *Proc. 16th Conference on Neural Information Processing Systems (NIPS)*, pp. 1401–1408.

Pasula, H., S. Russell, M. Ostland, and Y. Ritov (1999). Tracking many objects with many sensors. In *Proc. 16th International Joint Conference on Artificial Intelligence (IJCAI)*.

Patrick, D., J. Bush, and M. Chen (1973). Methods for measuring levels of well-being for a health status index. *Health Services Research 8*, 228–45.

Pearl, J. (1986a). A constraint-propagation approach to probabilistic reasoning. In *Proc. 2nd Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 357–370.

Pearl, J. (1986b). Fusion, propagation and structuring in belief networks. *Artificial Intelligence 29*(3), 241–88.

Pearl, J. (1987). Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence 32*, 245–257.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, California: Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*, 669–710.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press.

Pearl, J. and R. Dechter (1996). Identifying independencies in causal graphs with feedback. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 420–26.

Pearl, J. and A. Paz (1987). GRAPHOIDS: A graph-based logic for reasoning about relevance relations. In B. Du Boulay, D. Hogg, and L. Steels (Eds.), *Advances in Artificial Intelligence*, Volume 2, pp. 357–363. Amsterdam: North Holland.

Pearl, J. and T. S. Verma (1991). A theory of inferred causation. In *Proc. Conference on Knowledge Representation and Reasoning (KR)*, pp. 441–452.

Pe'er, D., A. Regev, G. Elidan, and N. Friedman (2001). Inferring subnetworks from preturbed expression profiles. *Bioinformatics 17*, S215–S224.

Peng, Y. and J. Reggia (1986). Plausibility of diagnostic hypotheses. In *Proc. 2nd Conference on Artificial Intelligence (AAAI)*, pp. 140–45.

Perkins, S., K. Lacker, and J. Theiler (2003, March). Grafting: Fast, incremental feature selection by gradient descent in function space. *Journal of Machine Learning Research 3*, 1333–1356.

Peterson, C. and J. R. Anderson (1987). A mean field theory learning algorithm for neural

networks. *Complex Systems 1*, 995–1019.

Pfeffer, A., D. Koller, B. Milch, and K. Takusagawa (1999). SPOOK: A system for probabilistic object-oriented knowledge representation. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 541–550.

Poh, K. and E. Horvitz (2003). Reasoning about the value of decision-model refinement: Methods and application. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 174–182.

Poland, W. (1994). *Decision Analysis with Continuous and Discrete Variables: A Mixture Distribution Approach*. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University.

Poole, D. (1989). Average-case analysis of a search algorithm for estimating prior and posterior probabilities in Bayesian networks with extreme probabilities. In *Proc. 13th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 606–612.

Poole, D. (1993a). Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence 64*(1), 81–129.

Poole, D. (1993b). The use of conflicts in searching Bayesian networks. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 359–367.

Poole, D. and N. Zhang (2003). Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research 18*, 263–313.

Poon, H. and P. Domingos (2007). Joint inference in information extraction. In *Proc. 23rd Conference on Artificial Intelligence (AAAI)*, pp. 913–918.

Pradhan, M. and P. Dagum (1996). Optimal Monte Carlo estimation of belief network inference. In *Proc. 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 446–453.

Pradhan, M., M. Henrion, G. Provan, B. Del Favero, and K. Huang (1996). The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence 85*, 363–97.

Pradhan, M., G. M. Provan, B. Middleton, and M. Henrion (1994). Knowledge engineering for large belief networks. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 484–490.

Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, New York.

Qi, R., N. Zhang, and D. Poole (1994). Solving asymmetric decision problems with influence diagrams. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 491–497.

Qi, Y., M. Szummer, and T. Minka (2005). Bayesian conditional random fields. In *Proc. 11thWork-shop on Artificial Intelligence and Statistics*.

Rabiner, L. R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE 77*(2), 257–286.

Rabiner, L. R. and B. H. Juang (1986, January). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 4–15.

Ramsey, F. (1931). *The Foundations of Mathematics and other Logical Essays*. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company. edited by R.B. Braithwaite.

Rasmussen, C. and C. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Rasmussen, C. E. (1999). The infinite gaussian mixture model. In *Proc. 13th Conference on Neural Information Processing Systems (NIPS)*, pp. 554–560.

Ravikumar, P. and J. Lafferty (2006). Quadratic programming relaxations for metric labelling and Markov random field MAP estimation. In *Proc. 23rd International Conference on Machine*

Learning (ICML).

Renooij, S. and L. van der Gaag (2002). From qualitative to quantitative probabilistic networks. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 422–429.

Richardson, M. and P. Domingos (2006). Markov logic networks. *Machine Learning 62*, 107–136.

Richardson, T. (1994). Properties of cyclic graphical models. Master's thesis, Carnegie Mellon University.

Riezler, S. and A. Vasserman (2004). Incremental feature selection and ll regularization for relaxed maximum-entropy modeling. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: John Wiley & Sons.

Rissanen, J. (1987). Stochastic complexity (with discussion). *Journal of the Royal Statistical Society, Series B 49*, 223–265.

Ristic, B., S. Arulampalam, and N. Gordon (2004). *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House Publishers.

Robert, C. and G. Casella (1996). Rao-Blackwellisation of sampling schemes. *Biometrika 83*(1), 81–94.

Robert, C. and G. Casella (2005). *Monte Carlo Statistical Methods* (2nd ed.). Springer Texts in Statistics.

Robins, J. M. and L. A. Wasserman (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 409–420.

Rose, D. (1970). Triangulated graphs and the elimination process. *Journal of Mathematical Analysis and Applications 32*, 597–609.

Ross, S. M. (1988). *A First Course in Probability* (third ed.). London: Macmillan.

Rother, C., S. Kumar, V. Kolmogorov, and A. Blake (2005). Digital tapestry. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. R. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Rusmevichientong, P. and B. Van Roy (2001). An analysis of belief propagation on the turbo decoding graph with Gaussian densities. *IEEE Transactions on Information Theory 48*(2).

Russell, S. and P. Norvig (2003). *Artificial Intelligence: A Modern Approach* (2 ed.). Prentice Hall.

Rustagi, J. (1976). *Variational Methods in Statistics*. New York: Academic Press.

Sachs, K., O. Perez, D. Pe'er, D. Lauffenburger, and G. Nolan (2005, April). Causal protein-signaling networks derived from multiparameter single-cell data. *Science 308*(5721), 523–529.

Sakurai, J. J. (1985). *Modern Quantum Mechanics*. Reading, Massachusetts: Addison-Wesley.

Santos, A. (1994). A linear constraint satisfaction approach to cost-based abduction. *Artificial Intelligence 65*(1), 1–28.

Santos, E. (1991). On the generation of alternative explanations with implications for belief revision. In *Proc. 7th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 339–347.

Saul, L., T. Jaakkola, and M. Jordan (1996). Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research 4*, 61–76.

Saul, L. and M. Jordan (1999). Mixed memory Markov models: Decomposing complex stochastic processes as mixture of simpler ones. *Machine Learning 37*(1), 75–87.

Saul, L. K. and M. I. Jordan (1996). Exploiting tractable substructures in intractable networks. In

Proc. 10th Conference on Neural Information Processing Systems (NIPS).

Savage, L. (1951). The theory of statistical decision. *Journal of the American Statistical Association 46*, 55–67.

Savage, L. J. (1954). *Foundations of Statistics*. New York: John Wiley & Sons.

Schäffer, A. (1996). Faster linkage analysis computations for pedigrees with loops or unused alleles. *Human Heredity*, 226–235.

Scharstein, D. and R. Szeliski (2003). High-accuracy stereo depth maps using structured light. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, Volume 1, pp. 195–202.

Schervish, M. (1995). *Theory of Statistics*. Springer-Verlag.

Schlesinger, M. (1976). Sintaksicheskiy analiz dvumernykh zritelnikh singnalov v usloviyakh pomekh (syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika 4*, 113–130.

Schlesinger, M. and V. Giginyak (2007a). Solution to structural recognition (max,+)-problems by their equivalent transformations (part 1). *Control Systems and Computers 1*, 3–15.

Schlesinger, M. and V. Giginyak (2007b). Solution to structural recognition (max,+)-problems by their equivalent transformations (part 2). *Control Systems and Computers 2*, 3–18.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Segal, E., D. Pe'er, A. Regev, D. Koller, and N. Friedman (2005, April). Learning module networks. *Journal of Machine Learning Research 6*, 557–588.

Segal, E., B. Taskar, A. Gasch, N. Friedman, and D. Koller (2001). Rich probabilistic models for gene expression. *Bioinformatics 17*(Suppl 1), S243–52.

Settimi, R. and J. Smith (2000). Geometry, moments and conditional independence trees with hidden variables. *Annals of Statistics*.

Settimi, R. and J. Q. Smith (1998a). On the geometry of Bayesian graphical models with hidden variables. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 472–479.

Settimi, R. and J. Q. Smith (1998b). On the geometry of Bayesian graphical models with hidden variables. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 472–479.

Shachter, R. (1988, July–August). Probabilistic inference and influence diagrams. *Operations Research 36*, 589–605.

Shachter, R. (1999). Efficient value of information computation. In *Proc. 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 594–601.

Shachter, R., S. K. Andersen, and P. Szolovits (1994). Global conditioning for probabilistic inference in belief networks. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 514–522.

Shachter, R. and D. Heckerman (1987). Thinking backwards for knowledge acquisition. *Artificial Intelligence Magazine 8*, 55 – 61.

Shachter, R. and C. Kenley (1989). Gaussian influence diagrams. *Management Science 35*, 527–550.

Shachter, R. and P. Ndilikilikesha (1993). Using influence diagrams for probabilistic inference and decision making. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 276–83.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research 34*, 871–882.

Shachter, R. D. (1989). Evidence absorption and propagation through evidence reversals. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 173–190.

Shachter, R. D. (1998). Bayes-ball: The rational pastime. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 480–487.

Shachter, R. D., B. D'Ambrosio, and B. A. Del Favero (1990). Symbolic probabilistic inference in belief networks. In *Proc. 6th Conference on Artificial Intelligence (AAAI)*, pp. 126–131.

Shachter, R. D. and M. A. Peot (1989). Simulation approaches to general probabilistic inference on belief networks. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 221–230.

Shachter, R. D. and M. A. Peot (1992). Decision making using probabilistic inference methods. In *Proc. 8th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 276–83.

Shafer, G. and J. Pearl (Eds.) (1990). *Readings in Uncertain Reasoning*. Representation and Reasoning. San Mateo, California: Morgan Kaufmann.

Shafer, G. and P. Shenoy (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence 2*, 327–352.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal 27*, 379–423; 623–656.

Shawe-Taylor, J. and N. Cristianini (2000). *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press.

Shenoy, P. (1989). A valuation-based language for expert systems. *International Journal of Approximate Reasoning 3*, 383–411.

Shenoy, P. (2000). Valuation network representation and solution of asymmetric decision problems. *European Journal of Operational Research 121*(3), 579–608.

Shenoy, P. and G. Shafer (1990). Axioms for probability and belief-function propagation. In *Proc. 6th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 169–198.

Shenoy, P. P. (1992). Valuation-based systems for Bayesian decision analysis. *Operations Research 40*, 463–484.

Shental, N., A. Zomet, T. Hertz, and Y. Weiss (2003). Learning and inferring image segmentations using the GBP typical cut algorithm. In *Proc. International Conference on Computer Vision*.

Shimony, S. (1991). Explanation, irrelevance and statistical independence. In *Proc. 7th Conference on Artificial Intelligence (AAAI)*.

Shimony, S. (1994). Finding MAPs for belief networks in NP-hard. *Artificial Intelligence 68*(2), 399–410.

Shoikhet, K. and D. Geiger (1997). A practical algorithm for finding optimal triangulations. In *Proc. 13th Conference on Artificial Intelligence (AAAI)*, pp. 185–190.

Shwe, M. and G. Cooper (1991). An empirical analysis of likelihood-weighting simulation on a large, multiply connected medical belief network. *Computers and Biomedical Research 24*, 453–475.

Shwe, M., B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. I. The probabilistic model and inference algorithms. *Methods of Information in Medicine 30*, 241–55.

Silander, T. and P. Myllymaki (2006). A simple approach for finding the globally optimal Bayesian network structure. In *Proc. 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*.

Singh, A. and A. Moore (2005). Finding optimal bayesian networks by dynamic programming. Technical report, Carnegie Mellon University.

Sipser, M. (2005). *Introduction to the Theory of Computation* (Second ed.). Course Technology.

Smith, A. and G. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B 55*, 3–23.

Smith, J. (1989). Influence diagrams for statistical modeling. *Annals of Statistics 17*(2), 654–72.

Smith, J., S. Holtzman, and J. Matheson (1993). Structuring conditional relationships in influence diagrams. *Operations Research 41*(2), 280–297.

Smyth, P., D. Heckerman, and M. Jordan (1997). Probabilistic independence networks for hidden Markov probability models. *Neural Computation 9*(2), 227–269.

Sontag, D. and T. Jaakkola (2007). New outer bounds on the marginal polytope. In *Proc. 21st Conference on Neural Information Processing Systems (NIPS)*.

Sontag, D., T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss (2008). Tightening LP relaxations for MAP using message passing. In *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Speed, T. and H. Kiiveri (1986). Gaussian Markov distributions over finite graphs. *The Annals of Statistics 14*(1), 138–150.

Spetzler, C. and C.-A. von Holstein (1975). Probabilistic encoding in decision analysis. *Management Science*, 340–358.

Spiegelhalter, D. and S. Lauritzen (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks 20*, 579–605.

Spiegelhalter, D. J., A. P. Dawid, S. L. Lauritzen, and R. G. Cowell (1993). Bayesian analysis in expert systems. *Statistical Science 8*, 219–283.

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proc. 11th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 491–98.

Spirtes, P., C. Glymour, and R. Scheines (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review 9*, 62–72.

Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Number 81 in Lecture Notes in Statistics. New York: Springer-Verlag.

Spirtes, P., C. Meek, and T. Richardson (1999). An algorithm for causal inference in the presence of latent variables and selection bias. See Glymour and Cooper (1999), pp. 211–52.

Srebro, N. (2001). Maximum likelihood bounded tree-width Markov networks. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Srinivas, S. (1993). A generalization of the noisy-or model. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 208–215.

Srinivas, S. (1994). A probabilistic approach to hierarchical model-based diagnosis. In *Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Studený, M. and R. Bouckaert (1998). On chain graph models for description of conditional independence structures. *Annals of Statistics 26*.

Sudderth, E., A. Ihler, W. Freeman, and A. Willsky (2003). Nonparametric belief propagation. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 605–612.

Sutton, C. and T. Minka (2006). Local training and belief propagation. Technical Report MSR-TR-2006-121, Microsoft Research.

Sutton, C. and A. McCallum (2004). Collective segmentation and labeling of distant entities in information extraction. In *ICML Workshop on Statistical Relational Learning and Its Connections to Other Fields*.

Sutton, C. and A. McCallum (2005). Piecewise training of undirected models. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*.

Sutton, C. and A. McCallum (2007). An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar (Eds.), *Introduction to Statistical Relational Learning*. MIT

Press.

Sutton, C., A. McCallum, and K. Rohanimanesh (2007, March). Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research 8*, 693–723.

Suzuki, J. (1993). A construction of Bayesian networks from databases based on an MDL scheme. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 266–273.

Swendsen, R. and J. Wang (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters 58*(2), 86–88.

Swendsen, R. H. and J.-S. Wang (1986, Nov). Replica Monte Carlo simulation of spin-glasses. *Physical Review Letters 57*(21), 2607–2609.

Szeliski, R., R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother (2008, June). A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence 30*(6), 1068–1080. See http://vision.middlebury.edu/MRF for more detailed results.

Szolovits, P. and S. Pauker (1992). Pedigree analysis for genetic counseling. In *Proceedings of the Seventh World Congress on Medical Informatics (MEDINFO '92)*, pp. 679–683. North-Holland.

Tanner, M. A. (1993). *Tools for Statistical Inference.* New York: Springer-Verlag.

Tarjan, R. and M. Yannakakis (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal of Computing 13*(3), 566–579.

Taskar, B., P. Abbeel, and D. Koller (2002). Discriminative probabilistic models for relational data. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 485–492.

Taskar, B., P. Abbeel, M.-F. Wong, and D. Koller (2007). Relational Markov networks. See Getoor and Taskar (2007).

Taskar, B., V. Chatalbashev, and D. Koller (2004). Learning associative Markov networks. In *Proc. 21st International Conference on Machine Learning (ICML)*.

Taskar, B., C. Guestrin, and D. Koller (2003). Max margin Markov networks. In *Proc. 17th Conference on Neural Information Processing Systems (NIPS)*.

Tatikonda, S. and M. Jordan (2002). Loopy belief propagation and Gibbs measures. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Tatman, J. A. and R. D. Shachter (1990). Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics 20*(2), 365–379.

Teh, Y. and M. Welling (2001). The unified propagation and scaling algorithm. In *Proc. 15th Conference on Neural Information Processing Systems (NIPS)*.

Teh, Y., M. Welling, S. Osindero, and G. Hinton (2003). Energy-based models for sparse over-complete representations. *Journal of Machine Learning Research 4*, 1235–1260. Special Issue on ICA.

Teyssier, M. and D. Koller (2005). Ordering-based search: A simple and effective algorithm for learning bayesian networks. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 584–590.

Thiele, T. (1880). *Sur la compensation de quelques erreurs quasisystematiques par la methode des moindres carrees.* Copenhagen: Reitzel.

Thiesson, B. (1995). Accelerated quantification of Bayesian networks with incomplete data. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-*

95), pp. 306–311. AAAI Press.

Thiesson, B., C. Meek, D. M. Chickering, and D. Heckerman (1998). Learning mixtures of Bayesian networks. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Thomas, A., D. Spiegelhalter, and W. Gilks (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 837–842. Oxford, UK: Clarendon Press.

Thrun, S., W. Burgard, and D. Fox (2005). *Probabilistic Robotics*. Cambridge, MA: MIT Press.

Thrun, S., D. Fox, W. Burgard, and F. Dellaert (2000). Robust Monte Carlo localization for mobile robots. *Artificial Intelligence 128*(1–2), 99–141.

Thrun, S., Y. Liu, D. Koller, A. Ng, Z. Ghahramani, and H. Durrant-Whyte (2004). Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research 23*(7/8).

Thrun, S., C. Martin, Y. Liu, D. Hähnel, R. Emery-Montemerlo, D. Chakrabarti, and W. Burgard (2004). A real-time expectation maximization algorithm for acquiring multi-planar maps of indoor environments with mobile robots. *IEEE Transactions on Robotics 20*(3), 433–443.

Thrun, S., M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot (2004). FastSLAM: An efficient solution to the simultaneous localization and mapping problem with unknown data association. *Journal of Machine Learning Research*.

Tian, J. and J. Pearl (2002). On the testable implications of causal models with hidden variables. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 519–527.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B 58*(1), 267–288.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 22*(4), 1701–1728.

Tong, S. and D. Koller (2001a). Active learning for parameter estimation in Bayesian networks. In *Proc. 15th Conference on Neural Information Processing Systems (NIPS)*, pp. 647–653.

Tong, S. and D. Koller (2001b). Active learning for structure in Bayesian networks. In *Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 863–869.

Torrance, G., W. Thomas, and D. Sackett (1972). A utility maximization model for evaluation of health care programs. *Health Services Research 7*, 118–133.

Tsochantaridis, I., T. Hofmann, T. Joachims, and Y. Altun (2004). Support vector machine learning for interdependent and structured output spaces. In *Proc. 21st International Conference on Machine Learning (ICML)*.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*, 1124–1131.

van der Merwe, R., A. Doucet, N. de Freitas, and E. Wan (2000a, Aug.). The unscented particle filter. Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department.

van der Merwe, R., A. Doucet, N. de Freitas, and E. Wan (2000b). The unscented particle filter. In *Proc. 14th Conference on Neural Information Processing Systems (NIPS)*.

Varga, R. (2000). *Matrix Iterative Analysis*. Springer-Verlag.

Verma, T. (1988). Causal networks: Semantics and expressiveness. In *Proc. 4th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 352–359.

Verma, T. and J. Pearl (1988). Causal networks: Semantics and expressiveness. In *Proc. 4th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 69–76.

Verma, T. and J. Pearl (1990). Equivalence and synthesis of causal models. In *Proc. 6th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 255 –269.

Verma, T. and J. Pearl (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In *Proc. 8th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 323–330.

Vickrey, D. and D. Koller (2002). Multi-agent algorithms for solving graphical games. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, pp. 345–351.

Vishwanathan, S., N. Schraudolph, M. Schmidt, and K. Murphy (2006). Accelerated training of conditional random fields with stochastic gradient methods. In *Proc. 23rd International Conference on Machine Learning (ICML)*, pp. 969–976.

Viterbi, A. (1967, April). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory 13*(2), 260–269.

von Neumann, J. and O. Morgenstern (1944). *Theory of games and economic behavior* (first ed.). Princeton, NJ: Princeton Univ. Press.

von Neumann, J. and O. Morgenstern (1947). *Theory of games and economic behavior* (second ed.). Princeton, NJ: Princeton Univ. Press.

von Winterfeldt, D. and W. Edwards (1986). *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press.

Vorobev, N. (1962). Consistent families of measures and their extensions. *Theory of Probability and Applications 7*, 147–63.

Wainwright, M. (2006). Estimating the "wrong" graphical model: Benefits in the computation-limited setting. *Journal of Machine Learning Research 7*, 1829–1859.

Wainwright, M., T. Jaakkola, and A. Willsky (2003a). Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Transactions on Information Theory 49*(5).

Wainwright, M., T. Jaakkola, and A. Willsky (2003b). Tree-reweighted belief propagation and approximate ML estimation by pseudo-moment matching. In *Proc. 9thWorkshop on Artificial Intelligence and Statistics*.

Wainwright, M., T. Jaakkola, and A. Willsky (2004, April). Tree consistency and bounds on the performance of the max-product algorithm and its generalizations. *Statistics and Computing 14*, 143–166.

Wainwright, M., T. Jaakkola, and A. Willsky (2005). MAP estimation via agreement on trees: Message-passing and linear programming. *IEEE Transactions on Information Theory*.

Wainwright, M., T. Jaakkola, and A. S. Willsky (2001). Tree-based reparameterization for approximate estimation on loopy graphs. In *Proc. 15th Conference on Neural Information Processing Systems (NIPS)*.

Wainwright, M., T. Jaakkola, and A. S. Willsky (2002a). Exact map estimates by (hyper)tree agreement. In *Proc. 16th Conference on Neural Information Processing Systems (NIPS)*.

Wainwright, M., T. Jaakkola, and A. S. Willsky (2002b). A new class of upper bounds on the log partition function. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Wainwright, M. and M. Jordan (2003). Graphical models, exponential families, and variational inference. Technical Report 649, Department of Statistics, University of California, Berkeley.

Wainwright, M. and M. Jordan (2004). Semidefinite relaxations for approximate inference on graphs with cycles. In *Proc. 18th Conference on Neural Information Processing Systems (NIPS)*.

Wainwright, M., P. Ravikumar, and J. Lafferty (2006). High-dimensional graphical model selection using ℓ_1-regularized logistic regression. In *Proc. 20th Conference on Neural Information*

Processing Systems (NIPS).

Warner, H., A. Toronto, L. Veasey, and R. Stephenson (1961). A mathematical approach to medical diagnosis — application to congenital heart disease. *Journal of the American Madical Association 177*, 177–184.

Weiss, Y. (1996). Interpreting images by propagating bayesian beliefs. In *Proc. 10th Conference on Neural Information Processing Systems (NIPS)*, pp. 908–914.

Weiss, Y. (2000). Correctness of local probability propagation in graphical models with loops. *Neural Computation 12*, 1–41.

Weiss, Y. (2001). Comparing the mean field method and belief propagation for approximate inference in MRFs. In M. Opper and D. Saad (Eds.), *Advanced mean field methods*, pp. 229–240. Cambridge, Massachusetts: MIT Press.

Weiss, Y. and W. Freeman (2001a). Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation 13*.

Weiss, Y. and W. Freeman (2001b). On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory 47*(2), 723–735.

Weiss, Y., C. Yanover, and T. Meltzer (2007). MAP estimation, linear programming and belief propagation with convex free energies. In *Proc. 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*.

Welling, M. (2004). On the choice of regions for generalized belief propagation. In *Proc. 20th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Welling, M., T. Minka, and Y. Teh (2005). Structured region graphs: Morphing EP into GBP. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI)*.

Welling, M. and S. Parise (2006a). Bayesian random fields: The Bethe-Laplace approximation. In *Proc. 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*.

Welling, M. and S. Parise (2006b). Structure learning in Markov random fields. In *Proc. 20th Conference on Neural Information Processing Systems (NIPS)*.

Welling, M. and Y.-W. Teh (2001). Belief optimization for binary networks: a stable alternative to loopy belief propagation. In *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*.

Wellman, M. (1985). Reasoning about preference models. Technical Report MIT/LCS/TR-340, Laboratory for Computer Science, MIT.

Wellman, M., J. Breese, and R. Goldman (1992). From knowledge bases to decision models. *Knowledge Engineering Review 7*(1), 35–53.

Wellman, M. and J. Doyle (1992). Modular utility representation for decision-theoretic planning. In *Procec. First International Conference on AI Planning Systems*, pp. 236–42. Morgan Kaufmann.

Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence 44*, 257–303.

Wellner, B., A. McCallum, F. Peng, and M. Hay (2004). An integrated, conditional model of information extraction and coreference with application to citation matching. In *Proc. 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 593–601.

Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *Journal of the American Statistical Association 75*, 963–975.

Werner, T. (2007). A linear programming approach to max-sum problem: A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence 29*(7), 1165–1179.

West, M. (1993). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Comput-*

ing Science and Statistics 24, 325–333.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester, United Kingdom: John Wiley and Sons.

Wiegerinck, W. (2000). Variational approximations between mean field theory and the junction tree algorithm. In *Proc. 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 626–636.

Wold, H. (1954). Causality and econometrics. *Econometrica 22*, 162–177.

Wood, F., T. Griffiths, and Z. Ghahramani (2006). A non-parametric bayesian method for inferring hidden causes. In *Proc. 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 536–543.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 557–85.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics 5*, 161–215.

Xing, E., M. Jordan, and S. Russell (2003). A generalized mean field algorithm for variational inference in exponential families. In *Proc. 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 583–591.

Yanover, C., T. Meltzer, and Y. Weiss (2006, September). Linear programming relaxations and belief propagation — an empirical study. *Journal of Machine Learning Research 7*, 1887–1907.

Yanover, C., O. Schueler-Furman, and Y. Weiss (2007). Minimizing and learning energy functions for side-chain prediction. In *Proc. International Conference on Research in Computational Molecular Biology (RECOMB)*, pp. 381–395.

Yanover, C. and Y. Weiss (2003). Finding the M most probable configurations using loopy belief propagation. In *Proc. 17th Conference on Neural Information Processing Systems (NIPS)*.

Yedidia, J., W. Freeman, and Y. Weiss (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Information Theory 51*, 2282–2312.

Yedidia, J. S., W. T. Freeman, and Y. Weiss (2000). Generalized belief propagation. In *Proc. 14th Conference on Neural Information Processing Systems (NIPS)*, pp. 689–695.

York, J. (1992). Use of the Gibbs sampler in expert systems. *Artificial Intelligence 56*, 115–130.

Yuille, A. L. (2002). CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation. *Neural Computation 14*, 1691–1722.

Zhang, N. (1998). Probabilistic inference in influence diagrams. In *Proc. 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 514–522.

Zhang, N. and D. Poole (1994). A simple approach to Bayesian network computations. In *Proceedings of the 10th Biennial Canadian Artificial Intelligence Conference*, pp. 171–178.

Zhang, N. and D. Poole (1996). Exploiting contextual independence in probabilistic inference. *Journal of Artificial Intelligence Research 5*, 301–328.

Zhang, N., R. Qi, and D. Poole (1993). Incremental computation of the value of perfect information in stepwise-decomposable influence diagrams. In *Proc. 9th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 400–407.

Zhang, N. L. (2004). Hierarchical latent class models for cluster analysis. *Journal of Machine Learning Research 5*, 697–723.

Zoeter, O. and T. Heskes (2006). Deterministic approximate inference techniques for conditionally Gaussian state space models. *Statistical Computing 16*, 279–292.

Zweig, G. and S. J. Russell (1998). Speech recognition with dynamic Bayesian networks. In *Proc. 14th Conference on Artificial Intelligence (AAAI)*, pp. 173–180.