Epilogue

It should be clear, particularly from the last chapter of this book, that the field of probabilistic networks and expert systems is developing rapidly, and no overview could hope to stay up-to-date for long. For this reason we have deliberately chosen to focus on a solid exposition of a limited class of issues, with some confidence that the fundamental methodology will remain of interest and continue to provide a basis for insights into the new challenges that will come about.

Such challenges are arising as the ideas associated with probabilistic expert systems spread far beyond their original focus on fixed networks of discrete variables with precisely specified probabilities. Real applications increasingly feature uncertainty about both quantitative and structural aspects of the model, possibly incomplete data, and a mixture of discrete and continuous nodes whose parent-child relationships may need to be expressed as statistical models. Dynamic networks, in which the structure changes in order to model some underlying evolving process, are a particularly important development whose application extends into the broader arena of signal processing and on-line monitoring.

The problems associated with such models are being tackled by a variety of professional disciplines, which makes the research area very exciting but also difficult to predict. Current developments suggest that the future will see an increasingly unified approach to complex stochastic systems that exploit conditional independence both for graphical representation and as a foundation for computation, and this perspective will incorporate artificial intelligence, signal processing, Bayesian statistics, machine learning, and a host of other disparate topics. By laying out a firm foundation for some of

the core ideas in probabilistic networks, we hope that we will have helped others to extend their application to currently unimagined horizons.

Appendix A
Conjugate Analysis for Discrete Data

A.1 Bernoulli process

Suppose that we observe a sequence of n independent Bernoulli trials, on each of which the probability of success is θ (and thus the probability of failure is $1-\theta$). The data d will consist of a sequence of successes and failures of length n. Denote by a, b, respectively, the numbers of successes and failures in d (thus $a+b=n$). Then the probability of observing the sequence d is

$$p(d \,|\, \theta) = \theta^a (1-\theta)^b. \tag{A.1}$$

If we do not keep the full information in d, but only the numbers a and b of successes and failures, we have instead (assuming that the total sample size was fixed in advance) the binomial formula

$$p(a, b \,|\, \theta) = \frac{(a+b)!}{a!\, b!}\, \theta^a (1-\theta)^b. \tag{A.2}$$

Alternatively, the data may have been collected by generating successive trials until exactly a successes had been observed. In this case (A.1) is still valid, but (A.2) is replaced by the negative binomial formula

$$p(a, b \,|\, \theta) = \frac{(a+b-1)!}{(a-1)!\, b!}\, \theta^a (1-\theta)^b. \tag{A.3}$$

Other more complex observation schemes can be envisaged. Suppose the observation scheme is *non-informative*, i.e., the reported sequence d incor-

porates all the data taken into account in deciding when to stop observation. Then (a, b) is a *sufficient statistic*, and the likelihood function for θ in any such case will again satisfy

$$L(\theta) \propto \theta^a (1-\theta)^b. \tag{A.4}$$

Suppose that the probability distribution representing prior uncertainty about the value of θ can taken to be the Beta distribution $\mathrm{Be}(\alpha, \beta)$, with density $b(\theta \,|\, \alpha, \beta)$ given by

$$b(\theta \,|\, \alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}. \tag{A.5}$$

The normalization constant, by which the right-hand side of (A.5) must be divided to give equality, is

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{A.6}$$

where $\Gamma(\cdot)$ denotes the standard gamma function. As a function of θ, the prior density has the same general mathematical form as the likelihood. The non-negative quantities α and β, determining the specific member of the Beta family used, are *hyperparameters*. With this prior the expected probability of success $\mathrm{E}(\theta)$ is given by $\alpha/(\alpha+\beta)$, and this also gives the marginal probability of success on a single trial. The variance of θ is given by

$$\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{A.7}$$

Combining the prior density (A.5) with the likelihood $L(\theta)$ from (A.4), using Bayes' theorem, yields the posterior density

$$p(\theta \,|\, d) \propto \theta^a(1-\theta)^b \times \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{a+\alpha-1}(1-\theta)^{b+\beta-1}. \tag{A.8}$$

That is, the posterior distribution of θ is $\mathrm{Be}(a+\alpha, b+\beta)$. We see that, if the prior is in the family of Beta distributions, then so too will be the posterior, based on any observations. We say that the Beta family is *conjugate* for sampling from the Bernoulli process.

For prediction, or for Bayesian model comparison (see Section 11.3), we need the marginal probability of the data, given by

$$p(d \,|\, \alpha, \beta) = \int p(d \,|\, \theta)\, b(\theta \,|\, \alpha, \beta)\, d\theta.$$

This is easily shown to yield

$$p(d \,|\, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+a)\Gamma(\beta+b)}{\Gamma(\alpha+\beta+a+b)}. \tag{A.9}$$

If, instead of the full data-sequence d, we want the marginal probability of observing the sufficient statistic values (a, b), we must multiply (A.9) by $(a+b)!/a!\,b!$ for binomial sampling, or by $(a+b-1)!/(a-1)!\,b!$ for negative binomial sampling. However, so long as we use the same data in all cases, such additional factors will not affect marginal likelihood comparisons among competing models.

A.2 Multinomial process

Now consider a discrete random variable X that can take on one of a set of k mutually exclusive and exhaustive values $\{x_1, \ldots, x_k\}$ with probabilities $\theta \equiv (\theta_1, \ldots, \theta_k)$, where $\theta_i > 0$, $\sum_i \theta_i = 1$. That is, $P(X = x_i \,|\, \theta) = \theta_i$, for $i = 1, \ldots, k$. For a sequence d of n independent observations distributed as X, of which n_1 are of type x_1, n_2 of type x_2, etc., the likelihood is given by

$$p(d \,|\, \theta) = \prod_{i=1}^{k} \theta_i^{n_i}, \tag{A.10}$$

and again a formula proportional to (A.10) will give the likelihood based on the sufficient statistic $(n_1, \ldots, n_k)$, under an arbitrary (non-informative) observation process. For example, if n is fixed in advance, we have the multinomial formula

$$p(n_1, \ldots, n_k \,|\, \theta) = \frac{n!}{\prod n_j!} \prod_j \theta_j^{n_j}. \tag{A.11}$$

Suppose that the probability distribution representing prior uncertainty about θ is the Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_k)$ having hyperparameter $\alpha \equiv (\alpha_1, \ldots, \alpha_k)$ with each $\alpha_i > 0$. That is, its density is

$$p(\theta) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}, \tag{A.12}$$

where we have introduced $\alpha_+ := \sum_i \alpha_i$. The marginal probability of observing $X = x_j$ is $\mathrm{E}(\theta_j) = \alpha_j/\alpha_+$.

The Dirichlet density again has a similar mathematical form to the likelihood, and thus the family of all Dirichlet distributions is conjugate for the multinomial sampling process. Posterior to observing data d (or the sufficient statistic $(n_1, \ldots, n_k)$), the revised distribution for θ will have density

$$p(\theta \,|\, d) \propto \prod_{i=1}^{k} \theta_i^{n_i + \alpha_i - 1}, \tag{A.13}$$

i.e., the posterior distribution is Dirichlet $\mathcal{D}(\alpha_1 + n_1, \ldots, \alpha_k + n_k)$. Again, the posterior is obtained by simple updating of each of the component hyperparameters, by adding on the number of cases observed to yield the associated outcome.

The marginal probability of the data (needed for example for purposes of prediction or model comparison) can be shown to be

$$p(d \mid \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_+)}{\prod_j \Gamma(\alpha_j)} \frac{\prod_j \Gamma(\alpha_j + n_j)}{\Gamma(\alpha_+ + n)}. \tag{A.14}$$

Under multinomial sampling with fixed total sample size n, the marginal probability of the sufficient statistic, i.e., of the counts $(n_1, \ldots, n_k)$, is obtained by multiplying (A.14) by $n!/\prod_j n_j!$.

Appendix B
Gibbs Sampling

The probabilistic networks considered in Chapters 6 and 7 allow exact local computation for the special cases of discrete domains, Gaussian domains, and mixed discrete-Gaussian domains. While useful for many applications, this imposes limitations of two different natures in applications. While quite large and complex networks of discrete random variables have been handled successfully, there can come a point at which the cardinality of the clique and separator potentials become too large to manipulate with current computers. An example of this arises in the analysis of time-series using dynamic graphs (Kjærulff 1995). A second limitation is that the distributions handled are too restrictive for many applications, and that more general distributions are desirable.

These sets of problems are approximately solved by the stochastic simulation technique known as Markov chain Monte Carlo (MCMC), which simulates samples from the required posterior distribution and uses these samples to estimate expectations of desired quantities. A particular form of MCMC is called *Gibbs sampling*, and can be shown to be especially appropriate for use on graphical models.

B.1 Gibbs sampling

Suppose that we have a discrete set of random variables $X = \{X_v : v \in V\}$, whose joint distribution P has *everywhere positive* density $p(x) > 0$.

Suppose further that we can find the *full conditional* distribution,

$$\mathcal{L}(X_v \,|\, X_{V\setminus\{v\}}), \tag{B.1}$$

for every variable. Then one can generate a sample from P as follows. First one initializes each of the n variables in X to some permissible configuration $(x_1^0, x_2^0, \ldots, x_n^0)$. Next, using (B.1) one samples in turn each individual variable X_v from its distribution conditional upon the current instantiation of the other variables, replacing the current instantiation of the variable with its sampled state. This process is repeated. Thus, at the (i,j)th stage we have a current instantiation $(x_1^j, x_2^j, \ldots, x_{i-1}^j, x_i^{j-1}, \ldots, x_n^{j-1})$. One then samples the ith variable from $\mathcal{L}(X_i \,|\, x_1^j, x_2^j, \ldots, x_{i-1}^j, x_{i+1}^{j-1}, \ldots, x_n^{j-1})$, to obtain $X_i = x_i^j$, say. It can be shown that, under broad regularity conditions, if this process is repeated a large number I of times one will obtain a configuration, $(X_1^I, \ldots, X_n^I)$ say, whose asymptotic distribution for $I \to \infty$ is P. In fact, one can be more general and not require the variables to be updated in sequence, as long as each is selected infinitely often. Each iteration, i.e., updating of all n variables, constitutes a transition of a Markov chain, whose equilibrium distribution is the joint distribution P.

If we could be confident that the Markov chain had converged to its equilibrium distribution by I iterations, then any desired expectation $\mathrm{E}f(X)$ could be estimated by the sample mean of the function f evaluated at a large number m of consecutive simulated values X^i:

$$\mathrm{E}f(X) \approx \frac{1}{m} \sum_{i=I+1}^{I+m} f(X^i), \tag{B.2}$$

and with probability equal to 1 this is an equality in the limit for $m \to \infty$.

How large I should be so that the asymptotic regime has been reached (i.e., the Markov chain has converged to its equilibrium distribution) is a topic of current research — see, for example, Cowles and Carlin (1996) for a review. The problem is that consecutive samples may be highly correlated, and so it may take a large number of iterations for a sample to 'forget' its initial values.

It is generally considered advisable to carry out a number of long runs starting from widely dispersed starting points (Gelman and Rubin 1992), whose samples can be pooled, once one is confident that the simulations have converged to a common distribution. Reducing the correlations between consecutive iterations (improving 'mixing') is also an important area of current research.

Notice the requirement that $p(x) > 0$. The Gibbs sampler will sometimes work if there are zeros in the probability density, but such zeros must not make one part of the state space become isolated from any other part by single variable changes. Thus, the presence of zeros must not generate constraints that act as barriers to prevent the sampler from roaming over

the whole of the possible configuration space. Technically, the resulting Markov chain must be *irreducible*.

An example of a reducible Markov chain is as follows. Suppose we have two variables A and B with a joint density $p(a, b)$ having zeros as shown in Figure B.1. Then, if we initialize the Gibbs sampler anywhere in the top left part, it will never be able to make a transition into the bottom right part, and vice versa. If the other entries are close to but not quite zero, then eventually it will (with probability equal to 1) make a transition from the top left to bottom right, but it may take a large number of iterations, of the order of the inverse of the probabilities. So in practice such 'near-irreducibility' can cause severe convergence problems. Irreducibility is a particular problem in models for genetic pedigrees, where one promising way of overcoming this type of problem has been to update large blocks of variables simultaneously in so-called *blocking Gibbs sampling* (Jensen et al. 1995).

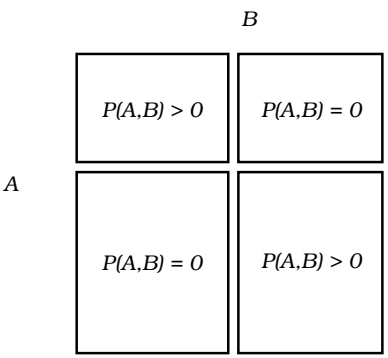

FIGURE B.1. Schematic illustrations of a probability density that can cause problems for a Gibbs sampler. Iterates starting in the top left region remain there and cannot migrate to the bottom right region because of the barrier generated by the zeros of the distribution, and vice versa. Hence, the Markov chain will be reducible.

B.2 Sampling from the moral graph

Having described the basic idea behind the Gibbs sampler, we now describe how it can be applied to probabilistic networks. Consider the simplest case of a probabilistic network consisting of a directed acyclic graph (DAG) of discrete random variables. The recursive factorization of the joint density

is given by (5.5):

$$p(x) = \prod_{v \in V} p(x_v \,|\, x_{\mathrm{pa}(v)}). \tag{B.3}$$

To apply the Gibbs sampler, we need to calculate for each node its distribution conditional upon all the other nodes. Now the only terms that involve a node v are the conditional density of X_v given its parents, and the conditional densities associated with each child of v. Taking only these terms yields a function which depends only on x_U, where $U = \{v\} \cup \mathrm{bl}(v)$ and $\mathrm{bl}(v)$ is the Markov blanket of v, defined as a node's parents, children, and co-parents, or equivalently its neighbours in the moral graph (see (5.6)). Hence, we can factorize $p(x)$ into two expressions, one containing all terms that involve x_v and the variables in its Markov blanket $\mathrm{bl}(v)$, and a second expression consisting of the remaining terms, which do not involve x_v and hence are irrelevant to the full conditional distribution. Hence,

$$\begin{aligned} p(x_v \,|\, x_{V \setminus \{v\}}) &= p(x_v \,|\, x_{\mathrm{bl}(v)}) \\ &\propto \text{terms in } p(x) \text{ containing } x_v \\ &= p(x_v \,|\, x_{\mathrm{pa}(v)}) \prod_{w \in \mathrm{ch}(v)} p(x_w \,|\, x_{\mathrm{pa}(w)}). \end{aligned} \tag{B.4}$$

Therefore, to calculate the full conditional distribution for a node v for the Gibbs sampler, one has only to consider the conditional distribution of X_v given its Markov blanket, which is easily calculated by (B.4). The same reasoning holds for undirected and chain graphs, in all cases the Markov blanket of a node being just its neighbours in the relevant moral graph.

Now suppose that one has some evidence $\mathcal{E}$ on some subset of variables. Then Gibbs sampling for the remaining nodes proceeds as above, but keeps the values of the observed nodes fixed throughout the simulation. Two caveats are: (1) the initial configuration must be permissible, and consistent with the evidence, and (2) the evidence must not make the resulting Markov chain reducible.

B.3 General probability densities

In principle, the theory of local computation via message-passing on a junction tree (Chapters 6 and 7) could be applied to probabilistic networks of more general types of random variables, with more complicated dependence of a variable on its parent configuration. The problem is that this generally leads to intractable integrals, so that the local computation cannot be performed in practice. However, this restriction does not apply to the Gibbs sampling procedure. Although the above description was based on discrete variables, it generalizes to arbitrary types of variables and dependencies.

Thus, if we are willing to give up exact local computation for approximate estimation based upon Gibbs sampling, then it is possible to treat much more general types of models. The factorization (B.4) still holds and so again we need only consider terms involving the Markov blanket. Sampling from a distribution proportional to the product of terms in (B.4) is not always straightforward, although a variety of special techniques has been developed.

These ideas form the basis of the BUGS (Bayesian Updating with Gibbs Sampling) computer program, which uses a specially designed high-level language to describe a graphical model (Gilks et al. 1994) and automatically constructs the necessary full conditional sampling procedures. WinBUGS is a prototype version that provides a graphical user interface both for specifying models and viewing results of the simulation. See Appendix C for the Web address.

B.4 Further reading

The use of Gibbs sampling in expert systems was first suggested by Pearl (1987). The PRESS system (Gammerman et al. 1995) incorporates options for both local propagation and Gibbs sampling for mixed discrete-Gaussian models.

Jensen et al. (1995) describe blocking Gibbs sampling, a variant of Gibbs sampling for expert system purposes, when the cliques of a junction tree are too large for practical computation. The basic idea is that variables are removed from the junction tree one at a time in order to reduce the size of cliques; this involves restructuring the junction tree. When the junction tree is of a computationally tractable size for exact local propagation, it is used to generate samples by the method described in Section 6.4.3, by a local exact computation. Thus, one samples clusters or *blocks* of variables simultaneously, so speeding up the convergence of the Gibbs sampler.

Dawid et al. (1995) have examined hybrid propagation in junction trees in which the natures of the potentials in the cliques and separators are heterogeneous and possibly change after receiving messages. Their scheme permits exact computation between some cliques and sampling in others.

The literature on the theory and application of Gibbs sampling is growing rapidly. For an exposition of recent developments see Gilks et al. (1996) and references therein; see also the Web addresses given in Appendix C. Recent surveys concerning convergence rates of Markov chains are Diaconis and Saloff-Coste (1995) and Jerrum and Sinclair (1996). Propp and Wilson (1996) showed that it is possible to sample exactly from the steady state distribution of certain types of Markov chain, using coupled Markov chains. Their work has stimulated much interest in algorithms of this type, called

exact sampling or *perfect sampling*. Appendix C gives the address of a Web page devoted to such techniques.

Appendix C

Information and Software on the World Wide Web

The field of probabilistic networks is developing rapidly, and most of the dissemination of non-academic information and software is via the World Wide Web. Below we provide a selection of sites that currently provide useful material and lists of links, so some exploration should rapidly reveal the current status of these sites and new ones that have been developed. A good starting point is the *Association for Uncertainty in Artificial Intelligence* site. Clearly, no recommendation or guarantee can be given as to the accuracy or suitability of any of the software mentioned.

The sites below were all functioning in April 1999. They are placed in alphabetical order of their URLs.

C.1 Information about probabilistic networks

`http://bayes.stat.washington.edu/almond/belief.html`

Russell Almond's page on software for manipulating belief networks:

Contains an extensive, but not necessarily up-to-date, list of both non-commercial and commercial software, with a glossary and reference list.

`http://http.cs.berkeley.edu/~murphyk/Bayes/bayes.html`

Kevin Murphy's introduction to Bayesian networks:

Contains a review, a good list of recommended reading with links to downloadable versions, and a page of free Bayesian network software (see below).

`http://www.afit.af.mil/Schools/EN/AI/`

US Air Force Institute of Technology Artificial Intelligence Laboratory:

Describes its PESKI software and has a lot of information and links on Bayesian networks. Included are a tutorial and summary of current research, while of particular interest are references to papers describing applications and reports from people in industry who have developed working systems.

`http://www.auai.org/`

Association for Uncertainty in Artificial Intelligence:

Contains links to proceedings of past Uncertainty in Artificial Intelligence conferences, as well as testimonials on the use of Bayesian networks, downloadable tutorials, and an excellent list of links to related sites of organizations, companies, and individuals.

`http://www.cs.auc.dk/research/DSS/`

Decision Support Systems Group at Aalborg University:

Describes their work on methodological development and practical application of Bayesian networks and influence diagrams.

`http://www.maths.nott.ac.uk/hsss/`

Highly Structured Stochastic Systems homepage:

Describes the background and aims of this initiative of the European Science Foundation, which brings together researchers who use stochastic models that exploit conditional independence. It puts research on probabilistic expert systems into a broader context, including Bayesian computation, genetics, environmental modelling, and image analysis.

`http://www.research.microsoft.com/research/dtg/`

Microsoft Research Decision Theory and Adaptive Systems group:

Gives brief description of projects being undertaken in information retrieval, diagnostics and troubleshooting, intelligent user interfaces, and so on, and how this work is being incorporated into Microsoft products. Some publications, particularly on learning models from data, are downloadable. Their free software, Microsoft Belief Networks (MSBN), is available (see next section).

C.2 Software for probabilistic networks

Just a few links are given here — see the above pages for a fuller and more up-to-date guide.

`http://hss.cmu.edu/html/departments/philosophy/TETRAD/tetrad.html`

The TETRAD project homepage:

Provides an overview and download information for the TETRAD software for building causal models from statistical data, and provides a publication list of project participants.

`http://http.cs.Berkeley.edu/~murphyk/Bayes/bnsoft.html`

Kevin Murphy's page of free Bayesian network software:

Gives links to a wide range of free packages, including his own Bayes Net Toolbox, and at the time of writing it is up-to-date.

`http://kmi.open.ac.uk/projects/bkd`

Bayesian Knowledge Discovery Project homepage:

Includes the free Bayesian Knowledge Discover program for automatically constructing discrete Bayesian networks from databases.

`http://www.city.ac.uk/~rgc`

Robert Cowell's home page:

Has a link to the freeware program XBAIES for building chain graph probabilistic and decision networks, available for several platforms.

`http://www.cs.cmu.edu/~javabayes/Home/`

Fabio Cozman's free JavaBayes program:

Includes download information and program details, with further links on Bayesian networks.

`http://www.math.auc.dk/~jhb/CoCo/cocoinfo.html`

Jens Henrik Badsberg's page for his COCO software:

Describes the software for graphical modelling of discrete variables, but is also a base for links and distribution of various other graphical modelling programs, including MIM, DIGRAM, BIFROST , and GAMES.

`http://www.mrc-bsu.cam.ac.uk/bugs/Welcome.html`

Home page of the BUGS project:

Includes the free WINBUGS program to build Bayesian graphical models. All inferences are by Markov chain Monte Carlo methods.

`http://www2.sis.pitt.edu/~genie/`

The GeNIe homepage:

A freeware program, developed by the Decisions System Laboratory at the University of Pittsburgh, for inference in Bayesian networks by a variety of methods. It can read and save files in most of the formats used by other similar programs. The Decisions System Laboratory also distributes a set of C++ classes, called SMILE, which programmers can use to build their own Bayesian network programs.

Commercial Bayesian network software includes:

- Hugin `http://www.hugin.dk/`
- DXpress `http://www.kic.com/`
- netica `http://www.norsys.com/netica.html`.

These pages contain numerous links to sites of interest.

C.3 Markov chain Monte Carlo methods

`http://dimacs.rutgers.edu/~dbwilson/exact.html`

This page contains a bibliography and pointers to papers concerned with perfect sampling from Markov chains.

`http://www.stats.bris.ac.uk/MCMC/`

Home page of a preprint service for papers concerned with Markov chain Monte Carlo methods.

Bibliography

Aczél, J. (1966). *Lectures on Functional Equations and their Applications*. Academic Press, New York.

Adams, I. D., Chan, M., Clifford, P. C., Cooke, W. M., Dallos, V., de Dombal, F. T., Edwards, M. H., Hancock, D. M., Hewett, D. J., McIntyre, N., Somerville, P. G., Spiegelhalter, D. J., Wellwood, J., and Wilson, D. H. (1986). Computer-aided diagnosis of acute abdominal pain: a multi-centre study. *British Medical Journal*, **292**, 800–4.

Aleliunas, R. (1990). A summary of a new normative theory of probabilistic logic. In *Uncertainty in Artificial Intelligence 4*, (ed. R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer), pp. 199–206. North-Holland, Amsterdam, The Netherlands.

Almond, R. (1995). *Graphical Belief Modeling*. Chapman and Hall, London, United Kingdon.

Andersen, K. A. and Hooker, J. N. (1994). Bayesian logic. *Decision Support Systems*, **11**, 191–210.

Andersen, L. R., Krebs, J. H., and Andersen, J. D. (1991). STENO: An expert system for medical diagnosis based on graphical models and model search. *Journal of Applied Statistics*, **18**, 139–53.

Andersen, S. K., Olesen, K. G., Jensen, F. V., and Jensen, F. (1989). HUGIN – A shell for building Bayesian belief universes for expert systems. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, (ed. N. S. Sridharan), pp. 1080–5. Morgan Kaufmann, San Mateo, California. Also reprinted in Shafer and Pearl (1990).

Andersson, S. A., Madigan, D., and Perlman, M. D. (1996a). An alternative Markov property for chain graphs. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. E. Horvitz and F. V. Jensen), pp. 40–8. Morgan Kaufmann, San Francisco, California.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1996b). A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics*, **25**, 505–41.

Andersson, S. A., Madigan, D., Perlman, M. D., and Richardson, T. (1998). Graphical Markov models in multivariate analysis. In *Multivariate Analysis, Design of Experiments and Survey Sampling*, (ed. S. Ghosh). Marcel Dekker Inc., New York.

Andreassen, S., Jensen, F. V., Andersen, S. K., Falck, B., Kjærulff, U., Woldbye, M., Sørensen, A., Rosenfalck, A., and Jensen, F. (1989). MUNIN — an expert EMG assistant. In *Computer-Aided Electromyography and Expert Systems*, (ed. J. E. Desmedt), pp. 255–77. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, The Netherlands.

Arnborg, S., Corneil, D. G., and Proskurowski, A. (1987). Complexity of finding embeddings in a k-tree. *SIAM Journal on Algebraic and Discrete Methods*, **8**, 277–84.

Bahl, L., Cocke, J., Jelinek, F., and Raviv, J. (1974). Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory*, **20**, 284–7.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley and Sons, New York.

Bartlett, M. S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society, Supplement*, **2**, 248–52.

Bauer, E., Koller, D., and Singer, Y. (1997). Update rules for parameter estimation in Bayesian networks. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. D. Geiger and P. Shenoy), pp. 3–13. Morgan Kaufmann, San Francisco, California.

Baum, L. E. (1972). An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, **3**, 1–8.

Becker, A. and Geiger, D. (1996). A sufficiently fast algorithm for finding close to optimal junction trees. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. E. Horvitz and F. V. Jensen), pp. 81–9. Morgan Kaufmann, San Francisco, California.

Beeri, C., Fagin, R., Maier, D., Mendelzon, A., Ullman, J., and Yannakakis, M. (1981). Properties of acyclic database schemes. In *Proceedings of the*

13th Annual ACM Symposium on the Theory of Computing, pp. 355–62. Association of Computing Machinery, New York.

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

Berge, C. (1973). *Graphs and Hypergraphs*. North-Holland, Amsterdam, The Netherlands. Translated from French by E. Minieka.

Bernardinelli, L., Pascutto, C., Best, N. G., and Gilks, W. R. (1997). Disease mapping with errors in covariates. *Statistics in Medicine*, **16**, 741–52.

Bernardo, J. M. (1979). Expected information as expected utility. *Annals of Statistics*, **7**, 686–90.

Bernardo, J. M. and Girón, F. J. (1988). A Bayesian analysis of simple mixture problems. In *Bayesian Statistics*, (ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith), pp. 67–78 (with discussion). Clarendon Press, Oxford, United Kingdom.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley and Sons, Chichester, United Kingdom.

Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistical Society, Series B*, **55**, 25–38.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.

Best, N. G., Spiegelhalter, D. J., Thomas, A., and Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, Series A*, **159**, 323–42.

Bielza, C. and Shenoy, P. P. (1997). A comparison of decision trees, influence diagrams and valuation networks for asymmetric decision problems. In *Preliminary Papers of the 6th International Workshop on Artificial Intelligence and Statistics*, pp. 39–46. Fort Lauderdale, Florida.

Bloemeke, M. and Valtorta, M. (1998). A hybrid algorithm to compute marginal and joint beliefs in Bayesian networks and its complexity. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 16–23. Morgan Kaufmann, San Francisco California.

Bollen, K. (1988). *Structural Equations with Latent Variables*. John Wiley and Sons, New York.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modelling and robustness (with discussion). *Journal of the Royal Statistical Society, Series A*, **143**, 383–430.

Box, G. E. P. (1983). An apology for ecumenism in statistics. In *Scientific Inference, Data Analysis and Robustness*, (ed. G. E. P. Box, T. Leonard, and C. F. Wu), pp. 51–84. Academic Press, New York.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, California.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**, 1–3.

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

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

Burnell, L. and Horvitz, E. J. (1995). Structure and chance: melding logic and probability for software debugging. *Communications of the ACM*, **38**, 31–41.

Call, H. J. and Miller, W. A. (1990). A comparison of approaches and implementations for automating decision analysis. *Reliability Engineering and System Safety*, **30**, 115–62.

Cannings, C., Thompson, E. A., and Skolnick, M. H. (1976). Recursive derivation of likelihoods on pedigrees of arbitrary complexity. *Advances in Applied Probability*, **8**, 622–5.

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

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago, Illinois.

Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1997). *Expert Systems and Probabilistic Network Models*. Springer–Verlag, New York.

Charnes, J. M. and Shenoy, P. P. (1997). A forward Monte Carlo method for solving influence diagrams using local computation. In *Preliminary Papers of the 6th International Workshop on Artificial Intelligence and Statistics*, pp. 75–82. Fort Lauderdale, Florida.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine*, **12**, 50–63.

Cheeseman, P. and Stutz, J. (1997). Bayesian classification (Autoclass): theory and results. In *Advances in Knowledge Discovery and Data Mining*, (ed. U. Fayyad, G. Pietesky-Shapiro, P. Smyth, and R. Uthurusamy), pp. 153–80. AAAI Press, Menlo Park, California.

Chen, T. and Fienberg, S. E. (1974). Two-dimensional contingency tables with both complete and partially cross-classified data. *Biometrics*, **30**, 629–42.

Chen, T. and Fienberg, S. E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics*, **32**, 133–44.

Chickering, D. M. (1996). Learning Bayesian networks is NP-complete. In *Learning from Data: Artificial Intelligence and Statistics V*, (ed. D. Fisher and H.-J. Lenz), pp. 121–30. Springer–Verlag, New York.

Chickering, D. M. and Heckerman, D. (1997). Efficient approximations for the marginal likelihood of incomplete data given a Bayesian network. *Machine Learning*, **29**, 181–212.

Chickering, D. M., Heckerman, D., and Meek, C. (1997). A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. D. Geiger and P. Shenoy), pp. 80–9. Morgan Kaufmann, San Francisco, California.

Chow, C. K. and Liu, C. N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions On Information Theory*, **14**, 462–7.

Christensen, R. (1990). *Log-Linear Models*. Springer–Verlag, New York.

Consonni, G. and Giudici, P. (1993). Learning in probabilistic expert systems. In *S.I.S. Workshop on Probabilistic Expert Systems, Roma, Oct. 14–15*, (ed. R. Scozzafava), pp. 57–78.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using belief networks. *Artificial Intelligence*, **42**, 393–405.

Cooper, G. F. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, **9**, 309–47.

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

Cowell, R. G. (1992). Calculating moments of decomposable functions in Bayesian networks. *Research Report* **109**, Department of Statistical Science, University College London, London, United Kingdom.

Cowell, R. G. (1994). Decision networks: a new formulation for multistage decision problems. *Research Report* **132**, Department of Statistical Science, University College London, London, United Kingdon.

Cowell, R. G. (1996). On compatible priors for Bayesian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 901–11.

Cowell, R. G. (1997). Sampling without replacement in junction trees. Technical Report **15**, Department of Actuarial Science and Statistics, City University London, London, United Kingdon.

Cowell, R. G. (1998). Mixture reduction via predictive scores. *Statistics and Computing*, **8**, 97–103.

Cowell, R. G. and Dawid, A. P. (1992). Fast retraction of evidence in a probabilistic expert system. *Statistics and Computing*, **2**, 37–40.

Cowell, R. G., Dawid, A. P., Hutchinson, T. A., Roden, S. R., and Spiegelhalter, D. J. (1993a). Bayesian networks for the analysis of drug safety. *The Statistician*, **42**, 369–84.

Cowell, R. G., Dawid, A. P., and Sebastiani, P. (1995). A comparison of sequential learning methods for incomplete data. In *Bayesian Statistics 5*, (ed. J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith), pp. 533–41. Clarendon Press, Oxford, United Kingdom.

Cowell, R. G., Dawid, A. P., and Spiegelhalter, D. J. (1993b). Sequential model criticism in probabilistic expert systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 209–19.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics — a comparative review. *Journal of the American Statistical Association*, **91**, 883–904.

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies*. Chapman and Hall, London, United Kingdom.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, **14**, 1–13.

Curds, R. M. (1997). *Propagation Techniques in Probabilistic Expert Systems*. Ph.D. Thesis, Department of Statistical Science, University College London.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics*, **8**, 522–39.

Dawid, A. P. (1976). Properties of diagnostic data distributions. *Biometrics*, **32**, 647–58.

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

Dawid, A. P. (1980a). A Bayesian look at nuisance parameters. In *Bayesian Statistics*, (ed. J. M. Bernardo, M. DeGroot, D. V. Lindley, and A. F. M. Smith), pp. 167–84. Valencia University Press, Valencia, Spain.

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

Dawid, A. P. (1982). Intersubjective statistical models. In *Exchangeability in Probability and Statistics*, (ed. G. Koch and F. Spizzichino), pp. 217–32. North-Holland, Amsterdam, The Netherlands.

Dawid, A. P. (1983). Invariant prior distributions. In *Encyclopedia of Statistical Sciences*, (ed. S. Kotz, N. L. Johnson, and C. B. Read), pp. 228–36. Wiley–Interscience, New York.

Dawid, A. P. (1984). Statistical theory. The prequential approach. *Journal of the Royal Statistical Society, Series A*, **147**, 277–305.

Dawid, A. P. (1985). Invariance and independence in multivariate distribution theory. *Journal of Multivariate Analysis*, **17**, 304–15.

Dawid, A. P. (1986). Probability forecasting. In *Encyclopedia of Statistical Sciences*, (ed. S. Kotz, N. L. Johnson, and C. B. Read), pp. 210–8. Wiley–Interscience, New York.

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

Dawid, A. P. (1992b). Prequential data analysis. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, IMS Lecture Notes–Monograph Series, **17**, (ed. M. Ghosh and P. K. Pathak), pp. 113–26. Institute of Mathematical Statistics, Hayward, California.

Dawid, A. P. (1998). Conditional independence. In *Encyclopedia of Statistical Sciences, Update Volume 2*, (ed. S. Kotz, C. B. Read, and D. L. Banks), pp. 146–55. Wiley–Interscience, New York.

Dawid, A. P. and Dickey, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *Journal of the American Statistical Association*, **72**, 845–50.

Dawid, A. P., Kjærulff, U., and Lauritzen, S. L. (1995). Hybrid propagation in junction trees. In *Advances in Intelligent Computing — IPMU 94*, Lecture Notes in Computer Science, **945**, (ed. B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh), pp. 87–97. Springer–Verlag, New York.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, **21**, 1272–317.

Dawid, A. P. and Skene, A. M. (1979). Maximum likelihood estimation of observer error rates using the EM algorithm. *Applied Statistics*, **28**, 20–8.

Dechter, R. (1996). Bucket elimination: A unifying framework for probabilistic inference. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. E. Horvitz and F. V. Jensen), pp. 211–9. Morgan Kaufmann, San Francisco, California.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw–Hill, New York.

Dempster, A. P. (1967). Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, **28**, 325–39.

Dempster, A. P., Laird, N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38 (with discussion).

Diaconis, P. and Saloff-Coste, L. (1995). What do we know about the Metropolis algorithm? In *Proceedings of the 27th Annual ACM Sym-*

posium on the Theory of Computing, pp. 112–29. Association of Computing Machinery, New York.

Didelez, V. and Pigeot, I. (1998). Maximum likelihood estimation in graphical models with missing values. *Biometrika*, **85**, 960–6.

Diestel, R. (1987). Simplicial decompositions of graphs – some uniqueness results. *Journal of Combinatorial Theory, Series B*, **42**, 133–45.

Diestel, R. (1990). *Graph Decompositions*. Clarendon Press, Oxford, United Kingdom.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.

Dirac, G. A. (1961). On rigid circuit graphs. *Abhandlungen Mathematisches Seminar Hamburg*, **25**, 71–6.

de Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., and Horrocks, J. C. (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, **2**, 9–13.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). *Biological Sequence Analysis*. Cambridge University Press, Cambridge, United Kingdom.

Edwards, D. (1990). Hierarchical interaction models (with discussion). *Journal of the Royal Statistical Society, Series B*, **52**, 3–20 and 51–72.

Edwards, D. (1995). *Introduction to Graphical Modelling*. Springer–Verlag, New York.

Edwards, D. and Havránek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, **72**, 339–51.

Edwards, D. and Havránek, T. (1987). A fast model selection procedure for large families of models. *Journal of the American Statistical Association*, **82**, 205–11.

de Finetti, B. (1937). Foresight: its logical laws, its subjective sources. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, **7**, 1–68. Translated by H. E. Kyburg in Kyburg and Smokler (1964), 55–118.

de Finetti, B. (1975). *Theory of Probability (Volumes 1 and 2)*. John Wiley and Sons, New York. Italian original 1970.

Franklin, R. C. G., Spiegelhalter, D. J., Macartney, F., and Bull, K. (1991). Evaluation of an algorithm for neonates. *British Medical Journal*, **302**, 935–9.

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

Friedman, N. (1998). The Bayesian structural EM algorithm. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 129–38. Morgan Kaufmann, San Francisco, California.

Friedman, N. and Goldszmidt, M. (1998). Learning Bayesian networks with local structure. In *Learning in Graphical Models*, (ed. M. I. Jordan), pp. 421–60. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**, 333–53.

Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, **77**, 270–8.

Fung, R. M. and Del Favero, B. (1995). Applying Bayesian networks to information retrieval. *Communications of the ACM*, **38**, 42–8.

Fung, R. M. and Shachter, R. D. (1990). Contingent influence diagrams. Working paper, Department of Engineering–Economic Systems, Stanford University.

Gabriel, K. R. (1969). Simultaneous test procedures — some theory of multiple comparisons. *Annals of Mathematical Statistics*, **40**, 224–50.

Gammerman, A., Luo, Z., Aitkin, C. G. G., and Brewer, M. J. (1995). Exact and approximate algorithms and their implementations in mixed graphical models. In *Probabilistic Reasoning and Bayesian Belief Networks*, (ed. A. Gammerman), pp. 33–53. Alfred Walker, Henley-on-Thames, United Kingdom.

Gavril, T. (1972). Algorithms for minimum coloring, maximum clique, minimum coloring by cliques and maximum independent set of a graph. *SIAM Journal on Computing*, **1**, 180–7.

Geiger, D. and Heckerman, D. (1994). Learning Gaussian networks. Technical Report MST-TR-94-10, Microsoft Corporation, Advanced Technology Division, Microsoft Corporation, Seattle, Washington.

Geiger, D., Heckerman, D., and Meek, C. (1998). Asymptotic model selection for directed networks with hidden variables. In *Learning in Graphical Models*, (ed. M. I. Jordan), pp. 461–78. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Geiger, D. and Meek, C. (1998). Graphical models and exponential families. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 156–65. Morgan Kaufmann, San Francisco, California.

Geiger, D. and Pearl, J. (1990). On the logic of causal models. In *Uncertainty in Artificial Intelligence 4*, (ed. R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer), pp. 3–14. North-Holland, Amsterdam, The Netherlands.

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

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

Gelman, A. and Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 625–31. Clarendon Press, Oxford, United Kingdom.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, **6**, 721–41.

Geng, Z., Asano, C., Ichimura, M., Tao, F., Wan, K., and Kuroda, M. (1996). Partial imputation method in the EM algorithm. In *Compstat 96*, (ed. A. Prat), pp. 259–63. Physica Verlag, Heidelberg, Germany.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo Methods in Practice*. Chapman and Hall, New York.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169–78.

Gillies, D. A. (1994). Philosophies of probability. In *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, (ed. I. Gratton-Guinness), pp. 1407–14. Routledge, London, United Kingdom.

Giudici, P. (1996). Learning in graphical Gaussian models. In *Bayesian Statistics 5*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 621–8. Clarendon Press, Oxford, United Kingdom.

Glymour, C., Scheines, R., Spirtes, P., and Kelley, K. (1987). *Discovering Causal Structure*. Academic Press, New York.

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

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, **14**, 107–14.

Goutis, C. (1995). A graphical method for solving a decision analysis problem. *IEEE Transactions on Systems, Man and Cybernetics*, **25**, 1181–93.

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

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, **82**, 711–32.

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

Harary, F. (1972). *Graph Theory*. Addison–Wesley, Reading, Massachusetts.

Heckerman, D. (1986). Probabilistic interpretations for MYCIN's certainty factors. In *Uncertainty in Artificial Intelligence*, (ed. L. N. Kanal and J. F. Lemmer), pp. 167–96. North-Holland, Amsterdam, The Netherlands.

Heckerman, D. (1990). Probabilistic similarity networks. *Networks*, **20**, 607–36.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In *Learning in Graphical Models*, (ed. M. I. Jordan), pp. 301–54. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Heckerman, D., Breese, J. S., and Rommelse, K. (1995a). Decision-theoretic troubleshooting. *Communications of the ACM*, **38**, 49–57.

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

Heckerman, D., Horvitz, E. J., and Nathwani, B. (1991). Toward normative expert systems I: The pathfinder project. *Methods of Information in Medicine*, **31**, 90–105.

Heckerman, D. and Wellman, M. P. (1995). Bayesian networks. *Communications of the ACM*, **38**, 27–30.

Henrion, M. (1988). Propagation of uncertainty by probabilistic logic sampling in Bayes' networks. In *Uncertainty in Artificial Intelligence 2*, (ed. J. F. Lemmer and L. N. Kanal), pp. 149–63. North-Holland, Amsterdam, The Netherlands.

Henrion, M., Breese, J. S., and Horvitz, E. J. (1991). Decision analysis and expert systems. *AI Magazine*, **12**, 64–91.

Hocking, R. R. and Oxspring, H. H. (1974). The analysis of partially categorized contingency data. *Biometrics*, **30**, 469–83.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, Massachusetts.

Howard, R. A. and Matheson, J. E. (1984). Influence diagrams. In *Readings in the Principles and Applications of Decision Analysis*, (ed. R. A. Howard and J. E. Matheson). Strategic Decisions Group, Menlo Park, California.

Højsgaard, S. and Thiesson, B. (1995). BIFROST — *B*lock recursive models *I*nduced *F*rom *R*elevant knowledge, *O*bservations and *S*tatistical *T*echniques. *Computational Statistics and Data Analysis*, **19**, 155–75.

Jeffreys, H. (1939). *Theory of Probability*. Clarendon Press, Oxford, United Kingdom.

Jensen, C. S., Kong, A., and Kjærulff, U. (1995). Blocking-Gibbs sampling in very large probabilistic expert systems. *International Journal of Human-Computer Studies*, **42**, 647–66.

Jensen, F. and Andersen, S. K. (1990). Approximations in Bayesian belief universes for knowledge-based systems. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, pp. 162–9. Cambridge, Massachusetts.

Jensen, F., Jensen, F. V., and Dittmer, S. L. (1994). From influence diagrams to junction trees. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, (ed. R. L. de Mantaras and D. Poole), pp. 367–73. Morgan Kaufmann, San Francisco, California.

Jensen, F. V. (1995). Cautious propagation in Bayesian networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, (ed. P. Besnard and S. Hanks), pp. 323–8. Morgan Kaufmann, San Francisco, California.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. University College London Press, London, United Kingdom.

Jensen, F. V. and Jensen, F. (1994). Optimal junction trees. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, (ed. R. L. de Mantaras and D. Poole), pp. 360–6. Morgan Kaufmann, San Francisco, California.

Jensen, F. V., Lauritzen, S. L., and Olesen, K. G. (1990a). Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, **4**, 269–82.

Jensen, F. V., Olesen, K. G., and Andersen, S. K. (1990b). An algebra of Bayesian belief universes for knowledge-based systems. *Networks*, **20**, 637–59.

Jerrum, M. and Sinclair, A. (1996). The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems*, (ed. D. S. Hochbaum). PWS Publishing, Boston, Massachusetts.

Jordan, M. I. (ed.) (1998). *Learning in Graphical Models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Kalman, R. E. and Bucy, R. (1961). New results in linear filtering and prediction. *Journal of Basic Engineering*, **83 D**, 95–108.

Kass, R. and Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association*, **90**, 773–95.

Kearns, M. and Saul, L. (1998). Large deviation methods for approximate probabilistic inference. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 311–9. Morgan Kaufmann, San Francisco, California.

Kellerer, H. G. (1964a). Maßtheoretische Marginalprobleme. *Mathematische Annalen*, **153**, 168–98.

Kellerer, H. G. (1964b). Verteilungsfunktionen mit gegebenen Marginalverteilungen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **3**, 247–70.

Keynes, J. M. (1921). *A Treatise on Probability.* Macmillan, London, United Kingdom.

Kim, J. H. and Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, (ed. A. Bundy), pp. 190–3. William Kaufmann, Los Altos, California.

Kjærulff, U. (1992). Optimal decomposition of probabilistic networks by simulated annealing. *Statistics and Computing*, **2**, 7–17.

Kjærulff, U. (1993). A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, (ed. D. Heckerman and A. Mamdani), pp. 121–9. Morgan Kaufmann, San Mateo, California.

Kjærulff, U. (1994). Reduction of computational complexity in Bayesian networks through removal of weak dependences. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, (ed. R. L. de Mantaras and D. Poole), pp. 374–82. Morgan Kaufmann, San Francisco, California.

Kjærulff, U. (1995). dHUGIN: a computational system for dynamic time-sliced Bayesian networks. *International Journal of Forecasting,* Special Issue on Probability Forecasting, **11**, 89–111.

Kjærulff, U. (1998). Nested junction trees. In *Learning in Graphical Models*, (ed. M. I. Jordan), pp. 51–74. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Koller, D. and Pfeffer, A. (1997). Object-oriented Bayesian networks. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. D. Geiger and P. Shenoy), pp. 302–13. Morgan Kaufmann, San Francisco, California.

Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability.* Chelsea Press, New York. German original 1933.

Kong, A. (1986). *Multivariate Belief Functions and Graphical Models.* Ph.D. Thesis, Department of Statistics, Harvard University, Massachusetts.

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

Kreiner, S. (1989). User's guide to DiGram — a program for discrete graphical modelling. Technical Report 89–10, Statistical Research Unit, University of Copenhagen.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Kyburg, H. E. and Smokler, H. E. (ed.) (1964). *Studies in Subjective Probability*. John Wiley and Sons, New York.

Lam, W. (1998). Bayesian network refinement via machine learning approach. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, **20**, 240–51.

Lam, W. and Bacchus, F. (1994). Learning belief networks: an approach based on the MDL principle. *Computational Intelligence*, **10**.

Larrañaga, P., Kuijpers, C. M. H., Poza, M., and Murga, R. H. (1997). Decomposing Bayesian networks: Triangulation of the moral graph with genetic algorithms. *Statistics and Computing*, **7**, 19–34.

Lauritzen, S. L. (1981). Time series analysis in 1880: A discussion of contributions made by T. N. Thiele. *International Statistical Review*, **49**, 319–31.

Lauritzen, S. L. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, **87**, 1098–108.

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. (1996). *Graphical Models*. Clarendon Press, Oxford, United Kingdom.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, **20**, 491–505.

Lauritzen, S. L. and Jensen, F. V. (1997). Local computation with valuations from a commutative semigroup. *Annals of Mathematics and Artificial Intelligence*, **21**, 51–69.

Lauritzen, S. L., Speed, T. P., and Vijayan, K. (1984). Decomposable graphs and hypergraphs. *Journal of the Australian Mathematical Society, Series A*, **36**, 12–29.

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

Lauritzen, S. L., Thiesson, B., and Spiegelhalter, D. J. (1994). Diagnostic systems created by model selection methods: A case study. In *Selecting Models from Data: AI and Statistics IV*, (ed. P. Cheeseman and R. Oldford), pp. 143–52. Springer–Verlag, New York. Lecture Notes in Statistics 89.

Lauritzen, S. L. and Wermuth, N. (1984). Mixed interaction models. Technical Report R 84-8, Institute for Electronic Systems, Aalborg University.

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

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Houghton-Mifflin, Boston, Massachusetts.

Leimer, H.-G. (1989). Triangulated graphs with marked vertices. *Annals of Discrete Mathematics*, **41**, 311–24.

Leimer, H.-G. (1993). Optimal decomposition by clique separators. *Discrete Mathematics*, **113**, 99–123.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, **44**, 187–92.

Lindley, D. V. (1982). Scoring rules and the inevitability of probability. *International Statistical Review*, **50**, 1–26.

Lindley, D. V. (1985). *Making Decisions*. John Wiley and Sons, Chichester, United Kingdom.

Lucas, P. and van der Gaag, L. (1991). *Principles of Expert Systems*. Addison–Wesley, Reading, Massachusetts.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, **89**, 1535–46.

Madigan, D. and York, I. (1994). Bayesian graphical models for discrete data. *International Statistical Review*, **63**, 215–32.

Madsen, A. L. and Jensen, F. V. (1998). Lazy propagation in junction trees. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 362–9. Morgan Kaufmann, San Francisco, California.

Maier, D. (1983). *The Theory of Relational Databases*. Computer Science Press, Rockville, Maryland.

McEliece, R. J., MacKay, D. J. C., and Cheng, J.-F. (1998). Turbo decoding as an instance of Pearl's belief propagation algorithm. *IEEE Journal on Selected Areas in Communications*, **16**, 140–52.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, (ed. P. Besnard and S. Hanks), pp. 403–10. Morgan Kaufmann, San Francisco, California.

Meilă, M. and Jordan, M. I. (1997). Triangulation by continuous embedding. In *Advances in Neural Information Processing Systems 9*, (ed. M. C. Mozer, M. I. Jordan, and T. Petsche), pp. 557–63. MIT Press, Cambridge, Massachusetts.

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

Monti, S. and Cooper, G. F. (1997). Learning Bayesian belief networks with neural network estimators. In *Advances in Neural Information Processing Systems 9*, (ed. M. C. Mozer, M. I. Jordan, and T. Petsche), pp. 579–84. MIT Press, Cambridge, Massachusetts.

Moussouris, J. (1974). Gibbs and Markov random systems with constraints. *Journal of Statistical Physics*, **10**, 11–33.

Murphy, A. H. and Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics*, **26**, 41–7.

Murphy, A. H. and Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association*, **79**, 489–500.

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

Neal, R. (1996). *Bayesian Learning for Neural Networks*. Springer–Verlag, New York.

Neapolitan, E. (1990). *Probabilistic Reasoning in Expert Systems*. John Wiley and Sons, New York.

Neyman, J. and Pearson, E. S. (1967). *Joint Statistical Papers of J. Neyman and E. S. Pearson*. University of California Press, Berkeley, California.

Nicholson, A. E. and Brady, J. M. (1994). Dynamic belief networks for discrete monitoring. *IEEE Transactions on Systems, Man and Cybernetics*, **24**, 1593–610.

Nilsson, D. (1994). An algorithm for finding the most probable configurations of discrete variables that are specified in probabilistic expert systems. M. Sc. Thesis, Department of Mathematical Statistics, University of Copenhagen, Denmark.

Nilsson, D. (1998). An efficient algorithm for finding the M most probable configurations in a probabilistic expert system. *Statistics and Computing*, **8**, 159–73.

Nilsson, N. J. (1986). Probabilistic logic. *Artificial Intelligence*, **28**, 71–87.

Normand, S.-L. and Tritchler, D. (1992). Parameter updating in a Bayes network. *Journal of the American Statistical Association*, **87**, 1109–15.

Olesen, K. G., Lauritzen, S. L., and Jensen, F. V. (1992). aHUGIN: A system creating adaptive causal probabilistic networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, (ed. D. Dubois, M. P. Wellman, B. D'Ambrosio, and P. Smets), pp. 223–9. Morgan Kaufmann, San Mateo, California.

Oliver, R. M. and Smith, J. Q. (1990). *Influence Diagrams, Belief Nets and Decision Analysis*. John Wiley and Sons, Chichester, United Kingdom.

Olmsted, S. M. (1983). *On Representing and Solving Decision Problems*. Ph.D. Thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, California.

Parter, S. (1961). The use of linear graphs in Gauss elimination. *SIAM Review*, **3**, 119–30.

Pearl, J. (1982). Reverend Bayes on inference engines: a distributed hierarchical approach. In *Proceedings of American Association for Artificial Intelligence National Conference on AI, Pittsburgh, Pennsylvania*, pp. 133–6.

Pearl, J. (1986a). A constraint–propagation approach to probabilistic reasoning. In *Uncertainty in Artificial Intelligence*, (ed. L. N. Kanal and J. F. Lemmer), pp. 357–70. North-Holland, Amsterdam, The Netherlands.

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

Pearl, J. (1987). Evidential reasoning using stochastic simulation. *Artificial Intelligence*, **32**, 245–57.

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

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

Pearl, J. and Paz, A. (1987). Graphoids: A graph based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence – II*, (ed. B. D. Boulay, D. Hogg, and L. Steel), pp. 357–63. North-Holland, Amsterdam, The Netherlands.

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

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London, United Kingdom. German original 1934.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–52.

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

Rabiner, L. R. and Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, New Jersey.

Raiffa, H. (1968). *Decision Analysis*. Addison–Wesley, Reading, Massachusetts.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. MIT Press, Cambridge, Massachusetts.

Ramoni, M. and Sebastiani, P. (1997a). Learning Bayesian networks from incomplete databases. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. D. Geiger and P. Shenoy), pp. 401–8. Morgan Kaufmann, San Francisco, California.

Ramoni, M. and Sebastiani, P. (1997b). Robust parameter learning in Bayesian networks with missing data. In *Preliminary Papers of the 6th International Workshop on Artificial Intelligence and Statistics*, pp. 339–406. Fort Lauderdale, Florida.

Ramoni, M. and Sebastiani, P. (1997c). The use of exogenous knowledge to learn Bayesian networks from incomplete databases. In *Advances in Intelligent Data Analysis. Reasoning about Data. Proceedings of the 2nd International Symposium on Intelligent Data Analysis Conference, London*, (ed. X. Liu, P. Cohen, and M. Berthold), pp. 537–48. Springer–Verlag, New York.

Ramsey, F. P. (1926). Truth and probability. In *The Foundation of Mathematics and Other Logical Essays*. Routledge and Kegan Paul, London, United Kingdom. Reprinted in Kyburg and Smokler (1964), pp. 61–92.

Richardson, T. (1996). A polynomial time algorithm for deciding Markov equivalence of directed cyclic graphical models. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. E. Horvitz and F. V. Jensen), pp. 462–9. Morgan Kaufmann, San Francisco, California.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, United Kingdom.

Rissanen, J. (1987). Stochastic complexity. *Journal of the Royal Statistical Society, Series B*, **49**, 223–39.

Riva, A. and Bellazzi, R. (1996). Learning temporal probabilistic causal-models from longitudinal data. *Artificial Intelligence in Medicine*, **8**, 217–34.

Robinson, R. W. (1977). Counting unlabelled acyclic digraphs. In *Lecture Notes in Mathematics: Combinatorial Mathematics V*, (ed. C. H. C. Little). Springer–Verlag, New York.

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

Rose, D. J. (1973). A graph theoretic study of the numerical solution of sparse positive definite systems of linear equation. In *Graph Theory and Computing*, (ed. R. Read), pp. 183–217. Academic Press, New York.

Rose, D. J., Tarjan, R. E., and Lueker, G. S. (1976). Algorithmic aspects of vertex elimination on graphs. *SIAM Journal on Computing*, **5**, 266–83.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–92.

Russell, S. J., Binder, J., Koller, D., and Kanazawa, K. (1995). Local learning in probabilistic networks with hidden variables. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, (ed. C. S. Mellish), pp. 1146–52. Morgan Kaufmann, San Mateo, California.

Sagiv, Y. and Walecka, S. F. (1982). Subset dependencies and a completeness result for a subclass of embedded multivalued dependencies. *Journal of the Association of Computing Machinery*, **29**, 103–17.

Sanguesa, R. and Cortes, U. (1997). Learning causal networks from data: a survey and a new algorithm for recovering possibilistic causal networks. *AI Communications*, **10**, 31–61.

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

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

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, **66**, 783–801.

Sebastiani, P. and Ramoni, M. (1998). Decision theoretic foundations of graphical model selection. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 464–71. Morgan Kaufmann, San Francisco, California.

Seillier-Moiseiwitsch, F. and Dawid, A. P. (1993). On testing the validity of sequential probability forecasts. *Journal of the American Statistical Association*, **88**, 355–9.

Seroussi, B. and Golmard, J. L. (1992). An algorithm for finding the K most probable configurations in Bayesian networks. *International Journal of Approximate Reasoning*, **1**, 205–33.

Settimi, R. and Smith, J. Q. (1998). On the geometry of Bayesian graphical models with hidden variables. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral). Morgan Kaufmann, San Francisco, California.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research*, **34**, 871–82.

Shachter, R. D. (1998). Bayes-ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 480–7. Morgan Kaufmann, San Francisco, California.

Shachter, R. D., Andersen, S. K., and Szolovits, P. (1994). Global conditioning for probabilistic inference in belief networks. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, (ed. R. L. de Mantaras and D. Poole), pp. 514–21. Morgan Kaufmann, San Francisco, California.

Shachter, R. D. and Kenley, C. (1989). Gaussian influence diagrams. *Management Science*, **35**, 527–50.

Shachter, R. D. and Ndilikilikesha, P. (1993). Using potential influence diagrams for probabilistic inference and decision making. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, (ed. D. Heckerman and A. Mamdani), pp. 383–90. Morgan Kaufmann, San Mateo, California.

Shachter, R. D. and Peot, M. A. (1990). Simulation approaches to general probabilistic inference on belief networks. In *Uncertainty in Artificial Intelligence 5*, (ed. M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer), pp. 221–31. North-Holland, Amsterdam, The Netherlands.

Shafer, G. R. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey.

Shafer, G. R. (1996). *Probabilistic Expert Systems*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

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

Shafer, G. R. and Shenoy, P. P. (1988). Local computation in hypertrees. Working Paper 201, School of Business, University of Kansas, Kansas.

Shafer, G. R. and Shenoy, P. P. (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence*, **2**, 327–52.

Sham, P. (1998). *Statistics in Human Genetics*. Edward Arnold, London, United Kingdon.

Shenoy, P. P. (1991). Valuation-based systems for discrete optimization. In *Uncertainty in Artificial Intelligence 6*, (ed. P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer), pp. 385–400. North-Holland, Amsterdam, The Netherlands.

Shenoy, P. P. (1992). Valuation-based systems for Bayesian decision analysis. *Operations Research*, **40**, 463–84.

Shenoy, P. P. (1994). A comparison of graphical techniques for decision analysis. *European Journal of Operational Research*, **78**, 1–21.

Shenoy, P. P. (1996). Representing and solving asymmetric decision problems using valuation networks. In *Learning from Data: Artificial Intelligence and Statistics V*, (ed. D. Fisher and H.-J. Lenz), pp. 99–108. Springer–Verlag, New York.

Shenoy, P. P. (1997). Binary join trees for computing marginals in the Shenoy–Shafer architecture. *International Journal of Approximate Reasoning*, **17**, 239–63.

Shenoy, P. P. and Shafer, G. R. (1986). Propagating belief functions using local propagation. *IEEE Expert*, **1**, 43–52.

Shenoy, P. P. and Shafer, G. R. (1990). Axioms for probability and belief–function propagation. In *Uncertainty in Artificial Intelligence 4*, (ed.

R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer), pp. 169–98. North-Holland, Amsterdam, The Netherlands.

Shortliffe, E. H. and Buchanan, B. G. (1975). A model for inexact reasoning in medicine. *Mathematical Biosciences*, **23**, 351–79.

Shwe, M., Middleton, B., Heckerman, D., Henrion, M., Horvitz, E. J., and Lehmann, H. (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.

Smith, A. F. M. and Makov, U. E. (1978). A quasi-Bayes sequential procedure for mixtures. *Journal of the Royal Statistical Society, Series B*, **40**, 106–11.

Smith, J. Q. (1979). A generalisation of the Bayesian steady forecasting model. *Journal of the Royal Statistical Society, Series B*, **41**, 375–87.

Smith, J. Q. (1981). The multiparameter steady model. *Journal of the Royal Statistical Society, Series B*, **43**, 256–60.

Smith, J. Q. (1989a). Influence diagrams for Bayesian decision analysis. *European Journal of Operational Research*, **40**, 363–76.

Smith, J. Q. (1989b). Influence diagrams for statistical modelling. *Annals of Statistics*, **7**, 654–72.

Smyth, P., Heckerman, D., and Jordan, M. I. (1997). Probabilistic independence networks for hidden Markov probability models. *Neural Computation*, **9**, 227–69.

Spiegelhalter, D. J. (1990). Fast algorithms for probabilistic reasoning in influence diagrams, with applications in genetics and expert systems. In *Influence Diagrams, Belief Nets and Decision Analysis*, (ed. R. M. Oliver and J. Q. Smith), pp. 361–84. John Wiley and Sons, Chichester, United Kingdom.

Spiegelhalter, D. J. and Cowell, R. G. (1992). Learning in probabilistic expert systems. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 447–65. Clarendon Press, Oxford, United Kingdom.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G. (1993). Bayesian analysis in expert systems (with discussion). *Statistical Science*, **8**, 219–83.

Spiegelhalter, D. J., Harris, N. L., Bull, K., and Franklin, R. C. G. (1994). Empirical evaluation of prior beliefs about frequencies: methodology and a case study in congenital heart disease. *Journal of the American Statistical Association*, **89**, 435–43.

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

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causality, Prediction and Search*. Springer–Verlag, New York.

Spirtes, P., Meek, C., and Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, (ed. P. Besnard and S. Hanks), pp. 499–506. Morgan Kaufmann, San Francisco, California.

Spirtes, P. and Richardson, T. (1997). A polynomial-time algorithm for determining DAG equivalence in the presence of latent variables and selection bias. In *Preliminary Papers of the 6th International Workshop on Artificial Intelligence and Statistics*, pp. 489–500. Fort Lauderdale, Florida.

Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In *Causation in Decision, Belief Change, and Statistics*, (ed. W. L. Harper and B. Skyrms), pp. 105–34. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society, Series B*, **36**, 111–47.

Studený, M. (1995). Conditional independence and natural conditional functions. *International Journal of Approximate Reasoning*, **12**, 934–45.

Studený, M. (1997). Comparison of graphical approaches to description of conditional independence structures. In *Proceedings of WUPES'97, January 22–25, 1997, Prague, Czech Republic*, pp. 156–72.

Studený, M. and Bouckaert, R. R. (1998). On chain graph models for description of conditional independence structures. *Annals of Statistics*, **26**, 1434–95.

Sy, B. K. (1993). A recurrence local computation approach towards ordering composite beliefs in Bayesian networks. *International Journal of Approximate Reasoning*, **8**, 17–50.

Tarjan, R. E. (1985). Decomposition by clique separators. *Discrete Mathematics*, **55**, 221–32.

Tarjan, R. E. and Yannakakis, M. (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing*, **13**, 566–79.

Tatman, J. A. and Shachter, R. D. (1990). Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics*, **20**, 365–79.

Thiele, T. N. (1880). Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en 'systematisk' Karakter. *Vidensk.*

Selsk. Skr. 5. Rk., naturvid. og mat. Afd., **12**, 381–408. French version: *Sur la Compensation de quelques Erreurs quasi-systématiques par la Méthode des moindres Carrés*. Reitzel, København, 1880.

Thiesson, B. (1991). (G)EM algorithms for maximum likelihood in recursive graphical association models. M. Sc. Thesis, Department of Mathematics and Computer Science, Aalborg University.

Thiesson, B. (1995). Accelerated quantification of Bayesian networks with incomplete data. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, (ed. U. M. Fayyad and R. Uthurusamy), pp. 306–11. AAAI Press, Menlo Park, California.

Thiesson, B. (1997). Score and information for recursive exponential models with incomplete data. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. D. Geiger and P. Shenoy), pp. 453–63. Morgan Kaufmann, San Francisco, California.

Thiesson, B., Meek, C., Chickering, D. M., and Heckerman, D. (1998). Learning mixtures of DAG models. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, (ed. G. F. Cooper and S. Moral), pp. 504–14. Morgan Kaufmann, San Francisco, California.

Titterington, D. M. (1976). Updating a diagnostic system using unconfirmed cases. *Applied Statistics*, **25**, 238–47.

Titterington, D. M., Murray, G. D., Murray, L. S., Spiegelhalter, D. J., Skene, A. M., Habbema, J. D. F., and Gelpke, G. J. (1981). Comparison of discrimination techniques applied to a complex data-set of head-injured patients (with discussion). *Journal of the Royal Statistical Society, Series A*, **144**, 145–75.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, Chichester, United Kingdom.

van der Gaag, L. (1991). Computing probability intervals under independency constraints. In *Uncertainty in Artificial Intelligence 6*, (ed. P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer), pp. 457–66. North-Holland, Amsterdam, The Netherlands.

Venn, J. (1866). *The Logic of Chance*. Macmillan, London, United Kingdom.

Verma, T. and Pearl, J. (1990). Causal networks: Semantics and expressiveness. In *Uncertainty in Artificial Intelligence 4*, (ed. R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer), pp. 69–76. North-Holland, Amsterdam, The Netherlands.

Verma, T. and Pearl, J. (1991). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence 6*, (ed. P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer), pp. 255–68. North-Holland, Amsterdam, The Netherlands.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**, 260–9.

von Mises, R. (1939). *Probability, Statistics and Truth*. Hodge, London, United Kingdom. German original 1928.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, New Jersey.

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

Vorob'ev, N. N. (1963). Markov measures and Markov extensions. *Theory of Probability and its Applications*, **8**, 420–9.

Vorob'ev, N. N. (1967). Coalition games. *Theory of Probability and its Applications*, **12**, 250–66.

Wagner, K. (1937). Über eine Eigenschaft der ebenen Komplexe. *Mathematische Annalen*, **114**, 570–90.

Walley, P. (1990). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, United Kingdom.

Warner, H. R., Toronto, A. F., Veasey, L. G., and Stephenson, R. (1961). A mathematical approach to medical diagnosis — application to congenital heart disease. *Journal of the American Medical Association*, **177**, 177–84.

Wermuth, N. (1976). Model search among multiplicative models. *Biometrics*, **32**, 253–63.

Wermuth, N. and Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of the Royal Statistical Society, Series B*, **52**, 21–72.

West, M. (1992). Modelling with mixtures. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 503–524 (with discussion). Clarendon Press, Oxford, United Kingdom.

West, M. and Harrison, P. J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer–Verlag, New York.

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

Winston, P. H. (1984). *Artificial Intelligence*, (2 edn). Addison–Wesley, Reading, Massachusetts.

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

Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics*, **8**, 239–55.

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

Yannakakis, M. (1981). Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic and Discrete Methods*, **2**, 77–9.

Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, **11**, 199–228.

Zhang, N. L. (1996). Irrelevance and parameter learning in Bayesian networks. *Artificial Intelligence*, **88**, 359–73.

Author Index

Subject Index