gRaphical models:

a softwaRe peRspective

Steffen L. Lauritzen, Aalborg University

Vienna, DSC 2003

Why this lecture?

- Graphical models have been around for about 25 years
- Software is the most important vehicle for dissemination of statistical ideas into practice
- Graphical models have shown some potential
- Software for graphical models exists as several independent stand-alone packages
- Time has come to attempt integration into general, flexible, and extendable software, such as R; hence gR.

Overview

- Brief historical sketch
- Basics of graphical models
- Present software situation
- Key dilemmas and challenges
- Basic needs and easy wins
- Demanding abstractions
- Where to go from here

Undirected graphical models

- Traces back to statistical physics (Gibbs 1902)
- Models for spatial interaction (Besag 1974)
- Interpreting hierarchical log-linear models by conditional independence, using analogy to Markov random fields (Darroch, Lauritzen and Speed 1980); CoCo (Badsberg 1991)
- Extending hierarchical log-linear models to include continuous variables (Lauritzen and Wermuth 1989); MIM (Edwards 1990).

Directed graphical models

- Traces back to path analysis (Wright 1921)
- Generalizing (block) recursive linear systems to discrete case (Wermuth and Lauritzen 1983, Wermuth and Lauritzen 1990); DIGRAM (Kreiner 1989).
- Bayesian (causal) networks (Pearl 1986, Lauritzen and Spiegelhalter 1988); HUGIN (Andersen, Olesen, Jensen and Jensen 1989), TETRAD (Spirtes, Glymour and Scheines 1993). DEAL (Bøttcher and Dethlefsen, DSC 2003).
- Bayesian graphical modelling; BUGS (Gilks, Thomas and Spiegelhalter 1994).

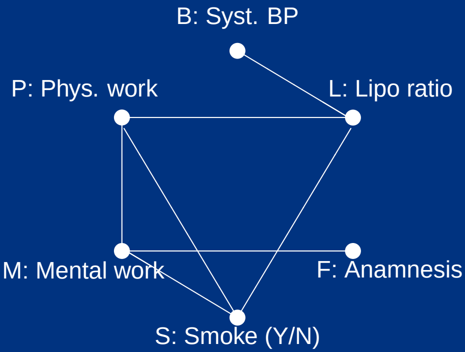
Undirected graphical models

- Nodes V represent set of variables $X_v, v \in V$
- Undirected graph $\mathcal{G} = (V, E)$ with (maximal) cliques $\mathcal{C}$.
- Joint density factorizes as

$$f(x) = \prod_{c \in \mathcal{C}} \phi_c(x),$$

  where ϕ_c depends on x through $x_c = (x_v)_{v \in c}$ only.
- Conditional independence: $A \perp\!\!\!\perp B \,|\, S$ if set of variables S separates A from B in $\mathcal{G}$.

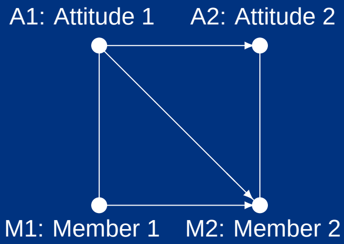
Undirected graphical models



$F \perp\!\!\!\perp S, B, P, L \mid M$ and $F, M \perp\!\!\!\perp L, B \mid S, P$ and $B \perp\!\!\!\perp S, P, M, F \mid L$.

show MIM

Directed (chain) graphical models



$$A2 \perp\!\!\!\perp M1 \mid A1, M2$$

show DIGRAM

Models for data

- Models shown so far are models for data.
- Graph represents model formula
- Parameters are implicit and not part of graph
- Data themselves are separate and not represented in graph
- Models are in essence multivariate
- Integration into R is ‘straightforward’ and fits with usual paradigm

Some simple advances

- Make interfaces between e.g. MIM and R (Højsgaard, DSC 2003)
- Make e.g. CoCo available within R (Badsberg, DSC 2003)
- Make R packages for special purposes e.g. DEAL (Bøttcher and Dethlefsen, DSC 2003)
- Make facilities directly within R so graphical modelling becomes easily extendable.

gRaphical models for data

Facilities to be desired within R to perform relevant analyses:

- Facilities for manipulating, representing, interacting with, displaying and printing graphs
- facilities and algorithms for specifying fitting, testing, selecting graphical models as in stand-alone programs
- Facilities for exporting and importing models and analyses to other programs

Bayesian networks

- Directed Acyclic Graph $\mathcal{D}$ (DAG)
- Nodes V represent variables $X_v, v \in V$
- Specify conditional distributions of children given parents: $p(x_v \,|\, x_{\mathrm{pa}(v)})$
- Joint distribution is then $p(x) = \prod_{v \in V} p(x_v \,|\, x_{\mathrm{pa}(v)})$
- Algorithm transforms network into *junction tree*
- Inference is performed by computing $p(x_v \,|\, x_A)$ which can be efficiently computed for all $A \subseteq V$.

show HUGIN

Data for models

BUGS extends idea of Bayesian networks for statistical purposes (complex Bayesian modelling)

- Parameters explicit, represented as nodes in graph
- Data explicitly represented in graph by observational nodes
- Special symbolism for repeated structures (plates)
- Inference by updating from prior to posterior, in effect calculating likelihoods by MCMC.

show BUGS

Compare simple linear regression

- Classical graphical model

- Standard model formula in R:

```
lm(y~x,data)
```

- BUGS model specification:

```
model{for(i in 1:N)
{Y[i]~ dnorm(mu[i], tau)
mu[i] <- a + b * (x[i] - mean(x[]))}
}
```

Contrasting paradigms

Bayesian vs. classical inference are not very different:

'Classical' graphical models can be treated with Bayesian methods

- Specify prior distributions
- Integrate likelihood functions rather than maximize
- Bayes factors instead of significance tests
- Posterior model probabilities for model search rather than comparison with significance tests, etc.

Asymptotically indistinguishable

Contrasting paradigms

Differences more radical concerning attitude towards modelling

- Data for models rather than models for data
- parameters, data, latent variables, covariates are fundamentally the same, i.e. random variables
- Status of variables changes dynamically through observation: X becomes x when observing $X = x$.
- Builds complex models using modularity and local modelling

Towards computing with models

- Local modelling paradigm is not Bayesian, but more fundamental
- Fits with modularity of graphical models
- In principle one could make LUGS, a likelihood version of BUGS. Just use prior distributions as 'convenient computational device'.
- More difficult to integrate local modelling paradigm into R. R has its origin within 'data analysis'; 'models' a later 'add on'.
- Needs further abstraction of data structures and object orientation.

Heads and Tails

Need local model object (LMO) $\mathrm{cd}(H \mid T)$ with head H and tail T.

Heads correspond to random variables, tails to parameters, in traditional thinking.

Variables in heads and tails can be instantiated by data. When T is empty or fully instantiated, $\mathrm{cd}(H \mid T)$ is a distribution, possibly represented as a program which simulates from it or otherwise can integrate.

When H is fully instantiated, $\mathrm{cd}(H \mid T)$ is a likelihood and is represented as a function which can be evaluated.

Operations with Models

With heads partially instantiated, LMOs represent conditional distribution of uninstantiated part of head given tail and instantiated part of head.

LMOs can be marginalised over uninstantiated parts of head. This reduces their head.

LMOs $\mathrm{cd}(H_1 \mid T_1)$ and $\mathrm{cd}(H_2 \mid T_2)$ can be combined when either $H_1 \cap (H_2 \cap T_2) = \emptyset$ or $H_2 \cap (H_1 \cap T_1) = \emptyset$ (or they satisfy specific consistency condition).

Graphs are formal objects which keep track of how LMOs are combined.

Where to go from here

- Continue first integration of stand-alones into R
- Identify specific elements of stand-alones which could usefully become part of gRaphical model library
- Identify new useful features which could usefully be added to such a graphical model library
- make repository of data for graphical models
- Do it!
- Continue process of abstraction to realise ‘computation with models’

References

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* N. S. Sridharan (ed.), *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, San Mateo, CA, pp. 1080–1085. Also reprinted in Shafer and Pearl (1990).

Badsberg, J. H.: 1991, A guide to CoCo, *Technical Report R-91-43*, Department of Mathematics and Computer Science, Aalborg University.

Besag, J.: 1974, Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B* **36**, 302–309.

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**, 522–539.

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

Gibbs, W.: 1902, *Elementary Principles of Statistical Mechanics*, Yale University Press, NewHaven, Connecticut.

Gilks, W. R., Thomas, A. and Spiegelhalter, D. J.: 1994, BUGS: a language and program for complex Bayesian modelling, *The Statistician* **43**, 169–178.

Kreiner, S.: 1989, User's guide to DIGRAM - a program for discrete graphical modelling, *Technical Report 89–10*, Statistical Research Unit, University of Copenhagen.

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

Pearl, J.: 1986, Fusion, propagation and structuring in belief networks, *Artificial Intelligence* **29**, 241–288.

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

Spirtes, P., Glymour, C. and Scheines, R.: 1993, *Causation, Prediction and Search*, Springer-Verlag, New York. Reprinted by MIT Press.

Wermuth, N. and Lauritzen, S. L.: 1983, Graphical and recursive models for contingency tables, *Biometrika* **70**, 537–552.

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.

Wright, S.: 1921, Correlation and causation, *Journal of Agricultural Research* **20**, 557–585.