rSMILE, an interface to the Bayesian Network package GeNIe/SMILE

Roman Klinger, Christoph M. Friedrich
{klinger,friedrich}@scai.fraunhofer.de
July 9, 2009

Outline

1 Bayesian Networks

2 Existing Implementations

3 GeNIe/SMILE

4 rSMILE

5 Applications

6 Acknowledgements

Full joint Distribution

- Given boolean variables
  Burglary, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Full joint distribution $P(b, e, a, j, m)$ has 2^5 combinations of input variables
  - $\Rightarrow$ Not intuitive
  - $\Rightarrow$ No causal interpretation

Stuart Russell and Peter Norvig
Artificial Intelligence – A Modern Approach.
Prentice Hall, 2003.

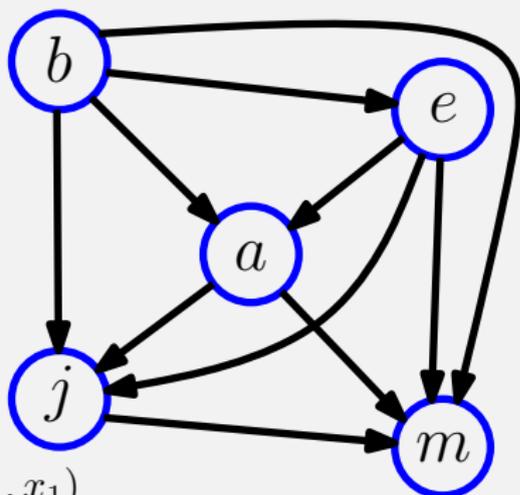# Bayesian Networks for full joint distribution

- Decomposition of a given probability distribution:

$$P(b, e, a, j, m) =$$
$$P(b)P(e|b)P(a|b, e)$$
$$P(j|b, e, a)P(m|b, e, a, j)$$
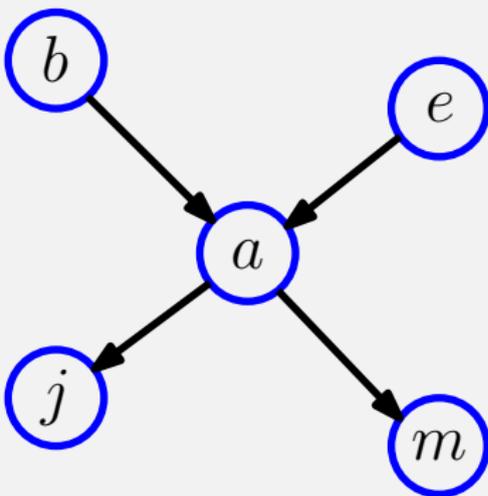
because of chain rule

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1)$$

Bayesian Networks with independence assumption

Assume independences:

- *Earthquake* independent of *Burglary*
- *JohnCalls*, *MaryCalls* independent of *Burglary*, *Earthquake*
- *MaryCalls* independent of *JohnCalls* given the *Alarm*

- Manually built network
- Simple, easy to interpret
- 10 numbers instead of 32

$$P(b, e, a, j, m) = P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

Bayesian Networks

Requirements for implementations

- Manual Generation of Network possible
- Learning of
    - Parameters
    - Structure
- Visualization Facilities
- Testing of single data points
- Interactive Use
- Evaluation of data sets
- Training and Evaluation on *nix machines
- High-performing Implementation

Existing Implementations

gR combining of several packages including

deal Learning Bayesian networks with mixed (discrete and continuous) variables

gRain Implements propagation in graphical models

BUGS Bayesian inference Using Gibbs Sampling

MIM,mimR Mixed Interaction Modeling - a Windows program for graphical modeling

TETRAD The TETRAD project: causal models and statistical data

📄 Søren Højsgaard

Graphical Models in R (gR)

`http://www.ci.tuwien.ac.at/gR/`

📄 Claus Dethlefsen and Søren Højsgaard

A Common Platform for Graphical Models in **R**: The g**R**base Package

Journal of Statistical Software, 14 (17), 2005

Fraunhofer
SCAI

Existing Approaches in R – BUGS

Bayesian inference Using Gibbs Sampling

- Graphical User Interface (winBUGS) for Windows (runs on Wine)

- Own Language ("Bugs Language")

- Usable in R ("BRugs")

- Complex, variety of versions available

📄 The BUGS Project
http://www.mrc-bsu.cam.ac.uk/bugs/
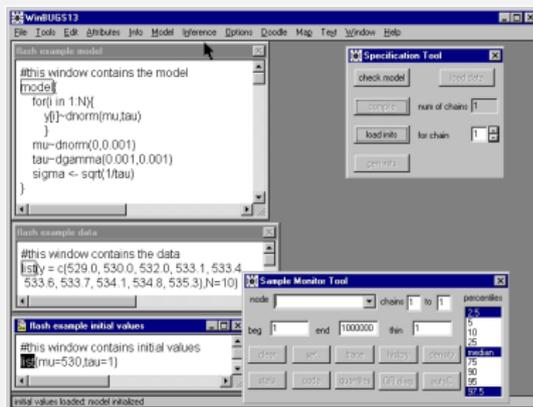
📄 OpenBUGS
http://mathstat.helsinki.fi/openbugs/

Existing Approaches in R – TETRAD

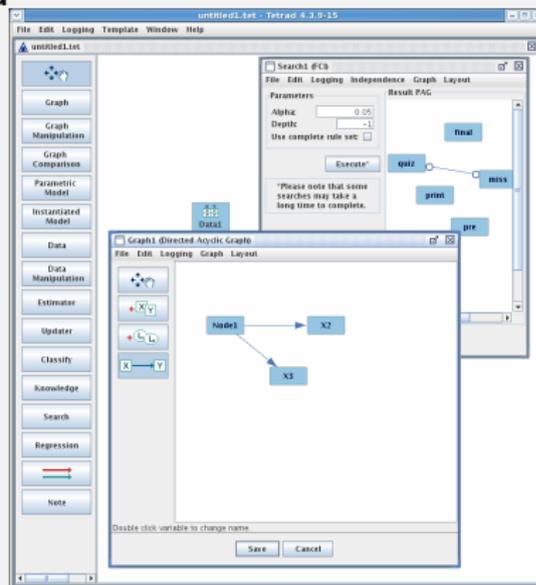## Causal Models and Statistical Data

- **Program for creating and testing in models**
- **Active development**
- **Pure Java**
- **Looks promising**
  (unknown to us at project time)
- **No R interface**

Peter Spirtes, Clark Glymour and Richard Scheines
The TETRAD Project
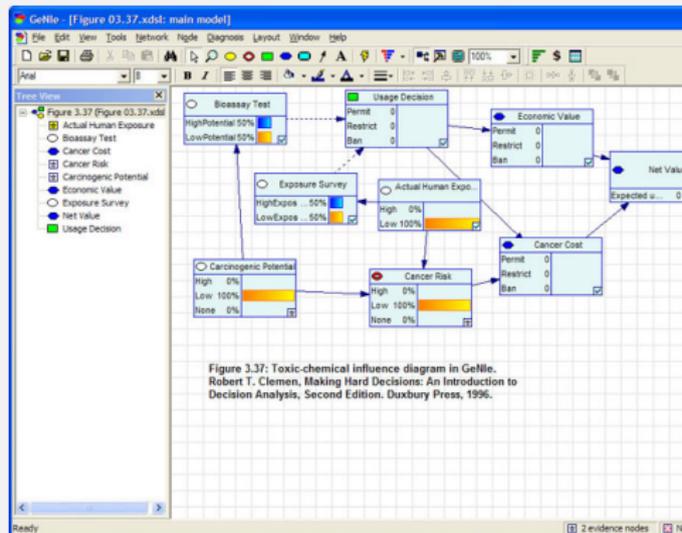http://www.phil.cmu.edu/projects/tetrad/

Fraunhofer
SCAI

GeNIe/SMILE

- **GeNIe** (Graphical Network Interface) as a user-friendly GUI (Windows, *nix with Wine)
- **SMILE** (Structural Modeling, Inference, and Learning Engine) as a cross-platform library
- Closed Source, but freely usable, even commercially
- Successfully applied in many publications

Figure 3.37: Toxic-chemical influence diagram in GeNIe.
Robert T. Clemen, Making Hard Decisions: An Introduction to
Decision Analysis, Second Edition. Duxbury Press, 1996.

Decision Systems Laboratory
GeNIe & SMILE, University of Pittsburgh
http://genie.sis.pitt.edu/

GeNIe/SMILE

- **GeNIe** (Graphical Network Interface) as a user-friendly GUI (Windows, *nix with Wine)

- **SMILE** (Structural Modeling, Inference, and Learning Engine) as a cross-platform library

- Closed Source, but freely usable, also commercially

- Successfully applied in many publications

Decision Systems Laboratory
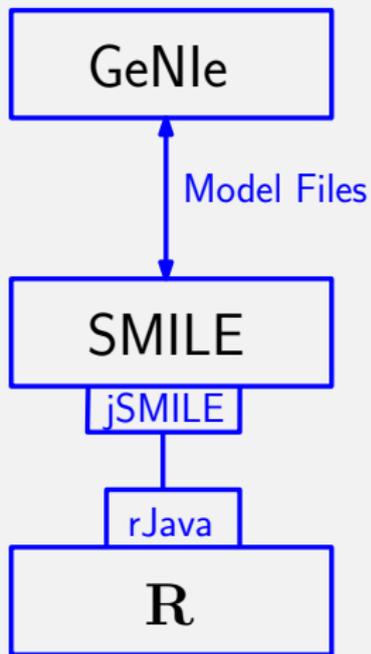
GeNIe & SMILE, University of Pittsburgh

http://genie.sis.pitt.edu/

+ Parameter learning algorithms

+ Structure learning algorithms
  + Background knowledge usable

+ Interactive use intuitive

+ High Performance

+ Automatic graph layout

− Only discrete variables

− Evaluation possibilities limited
  ⇒ Interface to **R**

rSMILE (1)

Interface from **R** to SMILE

- Based on jSMILE, the existing Java Native Interface to SMILE
- Based on rJava
- Features structure, parameter learning
- Evaluation possibilities
- Combines GeNIe's intuitive graphical interface with **R**'s comprehensive scriptability

rSMILE (2)

Features

- ■ Large networks processable (130 nodes tested)
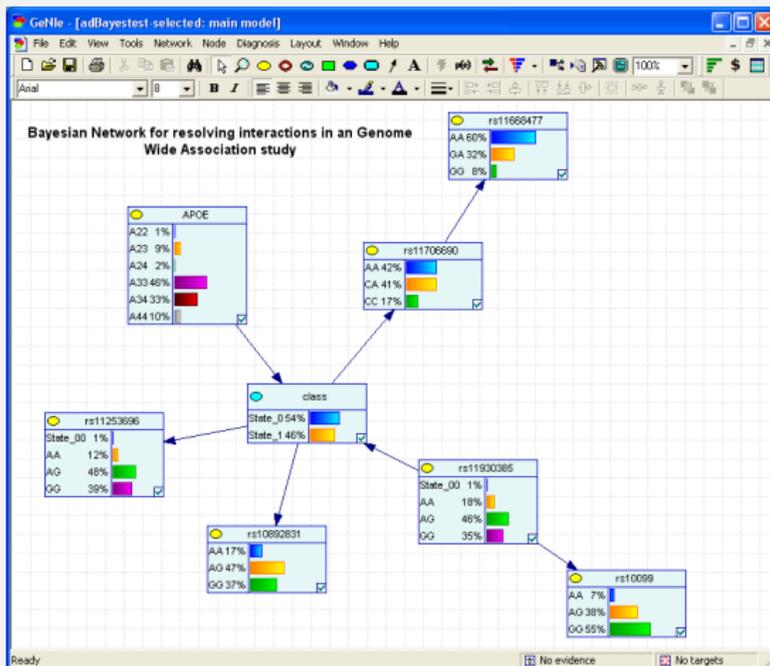- ■ Enhancement of SMILE-based Grow/Shrink algorithm to include background knowledge

Issues of rJava

- ■ Error/Exception Handling
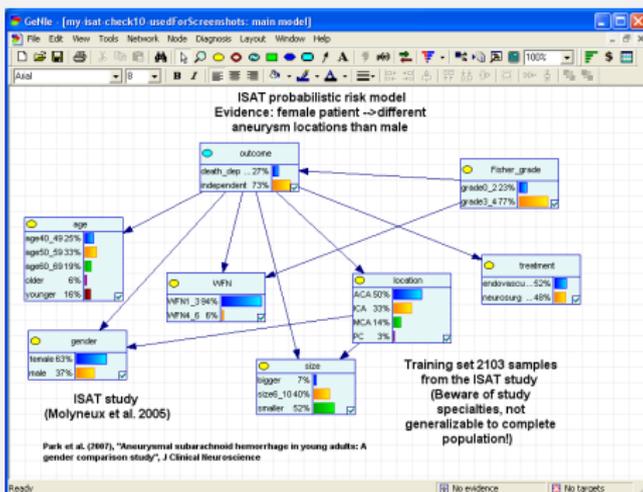- ■ Static parameters for JVM (first come $\rightarrow$ first serve)

Availability

- ■ Give us some time for code cleanup
  (until approx. September++)

Fraunhofer
SCAI

Interactions in Genome Wide Association Studies

International Study on Aneurysm Treatment

Molyneux, A. J.; Kerr, R. S. C.; Yu, L.; Clarke, M.; Sneade, M.; Yarnold, J. A. and Sandercock, P.
International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion
Lancet, 2005, 366, 809-817

⇒ Live Demo

Fraunhofer
SCAI

Acknowledgements

Thanks to

- Our co-workers: Martin Hofmann-Apitius, Beibei Han
- ISAT Consortium, Partners at @neurIST, especially Roelof Risselada
- This work has been partially funded in the framework of the European integrated project @neurIST, which is co-financed by the European Commission through the contract no. IST-027703 (see http://www.aneurist.org)

Thank YOU for your attention!

Questions?
Remarks?