Comparison between Suitable Priors for Additive Bayesian Networks

Gilles Kratzer[1], Reinhard Furrer[1,2], Marta Pittavino[3]

[1]Department of Mathematics, University of Zurich (Switzerland)

[2]Department of Computational Science, University of Zurich (Switzerland)

[3]Research Center for Statistics, Geneva School of Economy and Management, University of Geneva (Switzerland)

Contact: gilles.kratzer@math.uzh.ch

Motivation

- ABN methodology[1] extends the classical generalized linear model (GLM) framework to **multiple dependent variables**
- The key perspective of ABN is to extract the conditional independence information from an **observational dataset**
- ABN is a suitable methodology to mastermind **complex and messy data** in an exploratory analysis

Summary

- ABN is a mixture between **machine learning** techniques and a **statistical** approach
- ABN is distributed as an R package https://CRAN.R-project.org/package=abn
- **Gaussian, Binomial** and **Poisson** distributions are implemented in **abn**
- **Bayesian** based scoring function
- **Weakly informative** priors
- **Exact** and **Heuristic** search algorithm

Results

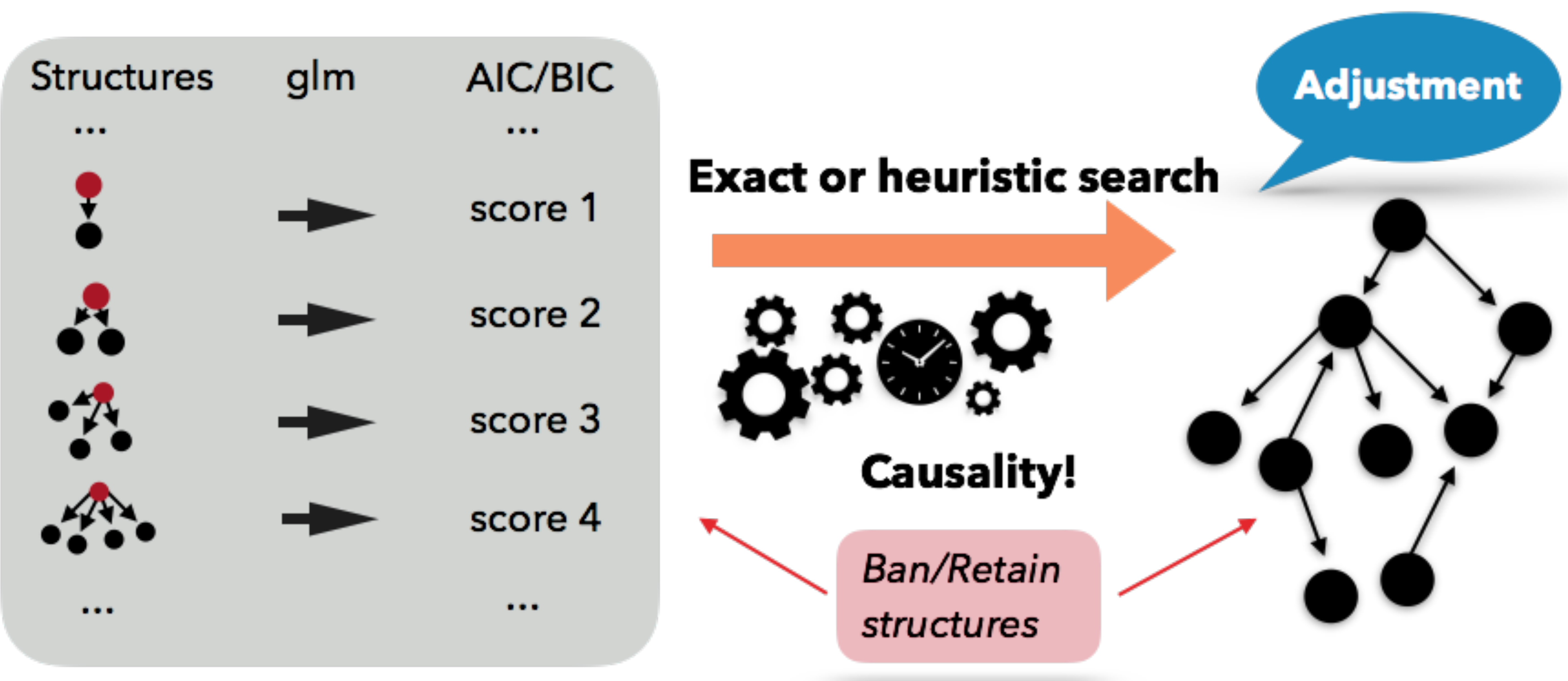
- Perform **structure discovery**
- ABN modelling empirically identifies associations in complex and high dimensional data as a **machine learning technique**

Future Work

- Implementation of informative priors suitable for ABN:

  Diaconis–Ylvisaker conjugate priors

How to fit an Additive Bayesian Network from data with the search-and-score method[2]?



Why priors matter in ABN[1]?

Data separation and data sparsity

- **Exclude** cases causing separation
- **Recast** the model:
  - **Remove** predictors
  - **Collapse** predictor categories
  - **Bin** predictors values
- Change a few randomly chosen observations
- Fit a hidden logistic regression model
- **Use priors to drive the posterior** when **marginal likelihood** is used as a proxy for BN relevance

Lyndley paradox

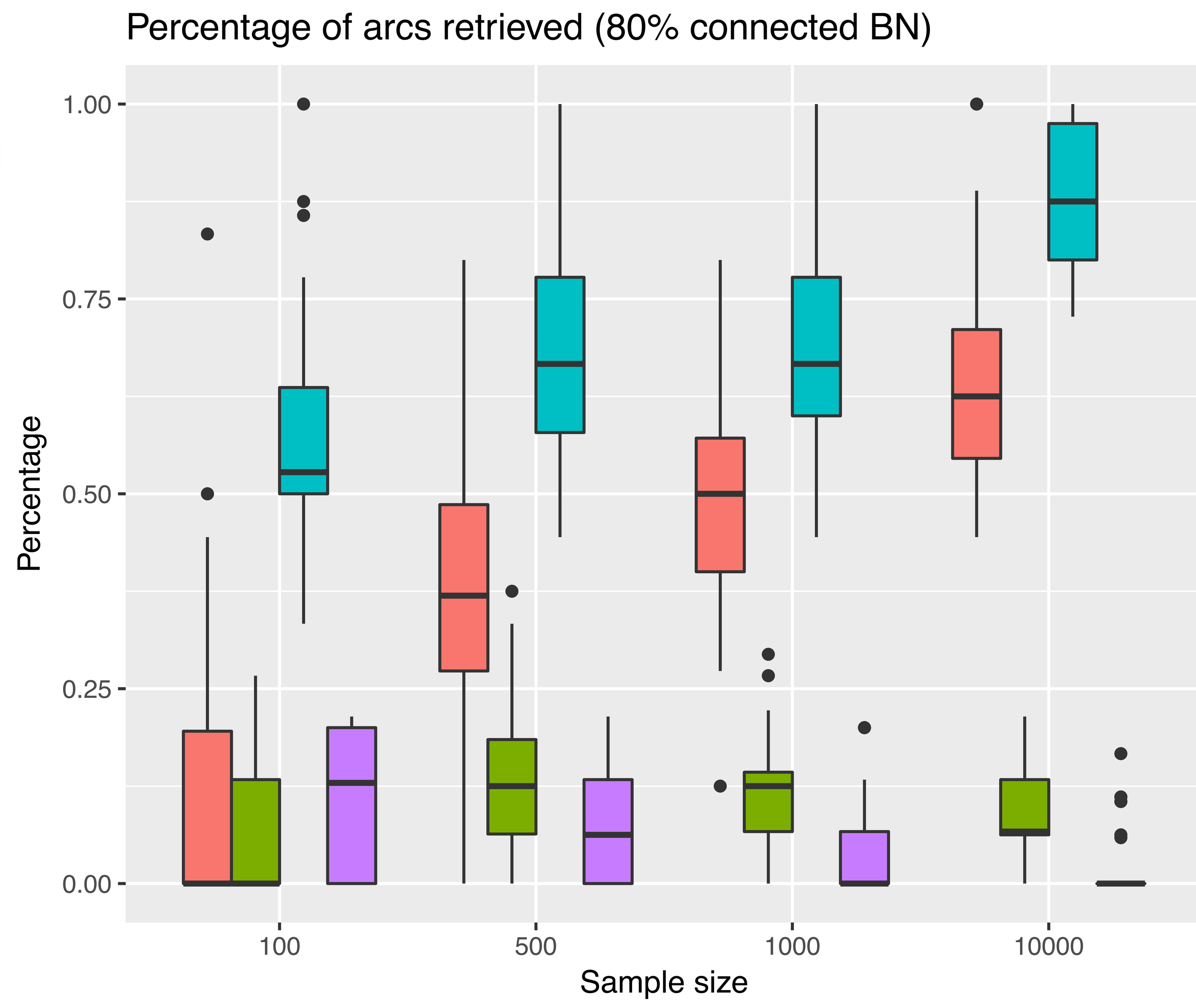
Bayesian model selection with nearly *uninformative* parameter priors tends to select simpler models **regardless** of the data

Simulation study

1. Simulate *many* different DAGs controlling for network density
2. Generate observations from the simulated DAGs using RNG
3. Fit ABN using different informative/weakly informative priors[3]
4. Compute DAGs metrics

Results

Data separation



expit(5) = 0.99

50 BN, n = 1000

Lyndley paradox

Student-t (ST) $\sim$ Cauchy$(0, 2.5)$

Weakly informative (WI) $\sim \mathcal{N}(0, 1000)$

Strongly informative (SI) $\sim \mathcal{N}(\theta_0, 0.1)$

expit(1) = 0.73

50 BN, n = 1000

References

1. Lewis, F. I. et al. "*Structure discovery in Bayesian networks: An analytical tool for analysing complex animal health data*", Preventive Veterinary Medicine 100.2 (2011): 109-115
2. Koivisto, M. et al.."*Exact Bayesian structure discovery in Bayesian networks*" Journal of Machine Learning Research, (2004) 549-573
3. Gelman, A, et al. "*A weakly informative default prior distribution for logistic and other regression models.*" The Annals of Applied Statistics 2.4 (2008): 1360-1383