Optimal independence tests for Bayesian Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 Introduction

Learning Bayesian networks for general distributions is intractable task Chickering [1996]. However, often for real data distributions, still we are able to recover Bayesian network structure. This implies that real data distribution has some special properties, which simplify process of structure recovery. Such process is almost always based on computation local statistics, and then reasoning about global structure Jaakkola et al. [2010]; Tsamardinos et al. [2006].

Such local statistics can describe complexity of network (e.g. number of parameters in case of BICSchwarz et al. [1978]), or can measure dependency of between nodes (e.g. mutual information tests, conditional independence tests). We focus in this work on how to best measure dependency between nodes in empirical CPDs from Bayessian network. We are mainly interest in developing techniques applicable to discovery of structure in gene expression data. This implies that each random variable can belong to many classes (when expression is quantized), or have continuous values. Moreover, such datasets are always relatively small ~ 200 samples. We focus our attention on addressing Bayesian network structure learning in aforementioned regime.

Our main contribution is in designing very efficient dependency test for Bayesian networks. It is kernelized partial correlation with kernels sensitive to linear dependency present in CPDs. We show that kernelized partial correlation with proper kernel in "gene expression" regime is almost always better than any other previously considered dependency test.

2 Related work

Independence tests used in Bayesian networks

- Schäfer and Strimmer [2005] Here they learn a Gaussian Graphical Model (GGM) using estimates of the partial correlation matrix.
- Opgen-Rhein and Strimmer [2007] Here they learn an approximate causal structure on gene expression based on full-order partial correlation (as an approximation to lower-order partial correlation that is called for theoretically for a Bayesian network).
- Tsamardinos et al. [2006] MMHC algorithm. They use a test based on what is called the G^2 statistic (asymptotically distributed as chi^2) and they also talk about a couple other independence tests that we may want to look into.

There has been extensive research in area of scoring functions for Bayesian networks. This step is crucial to prevent algorithms from recovering complete graph structure. Without any scoring function, log-likelihood term would force optimization to choose fully connected graph. There have been proposed few regularizations (scoring functions) to address this problem. There are two general types of scoring functions (1) based purely on model complexity, (2) combining model complexity with data evidences.

To the first group belongs one most widely used scoring functions, which is Bayesian Information Criterion (BIC) Schwarz et al. [1978]. There are couple of others like HannanQuinn information criterion (HQC) Hannan and Quinn [1979], or Bayesian model comparison (BMC). Major drawback of scoring functions based purely on model complexity is their constant power regardless of amount of data.

Variety of such scoring functions calculate dependency between nodes conditioned on potential parents De Campos [2006]. Usual measures of dependency are based on mutual information, conditional independence test, or are fully Bayesian. Fully Bayesian methods assume probability distribution over CPDs of independent variables, and dependent variables.

We should discuss:

- LL (Log-likelihood) (1912-22)
- MDL/BIC (Minimum description length/Bayesian Information Criterion) (1978)
- AIC (Akaike Information Criterion) (1974)
- NML (Normalized Minimum Likelihood) (2008)
- MIT (Mutual Information Tests) (2006)

Use in Bayesian networks Schäfer and Strimmer [2005]

Existing independence tests:

- Pearson's χ-squared. The problem is the null hypothesis is independence, but independence is what we're trying to show.

Margaritis [2003]

3 Independence testing

TODO : Wojciech (give a pass to Arthur Gretton) Independence tests have to decide if random variables are independent. Samples of this random variables gives us indirect access to the conditional probability distribution (CPD), which decides on independence. However, samples itself provide only empirical estimate on conditional probability distribution. Moreover, manifold 1 of independent CPDs among all possible CPDs have a measure zero.

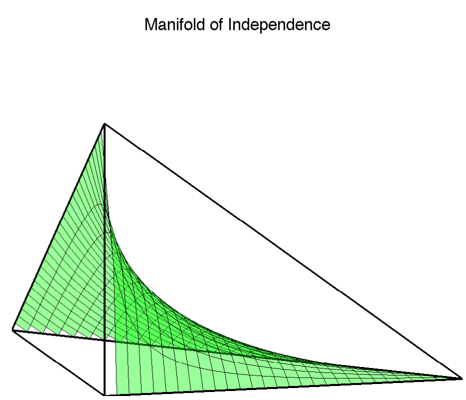

Figure 1: The manifold of independence for binary distributions. The simplex represents all possible joint distributions over two binary variables. a, b, and c are three of the four entries in the joint distribution table, and the simplex is formed by the constraint that all entries must be positive and sum to one. The manifold corresponds to the set (of measure zero) of independent distributions.

3.1 Curse of conditioning

3.2 Partial correlation

3.3 Kernelized partial correlation

4 Experiments

4.1 Classification of Synthetic CPDs

We generated distributions from toy Bayesian networks.

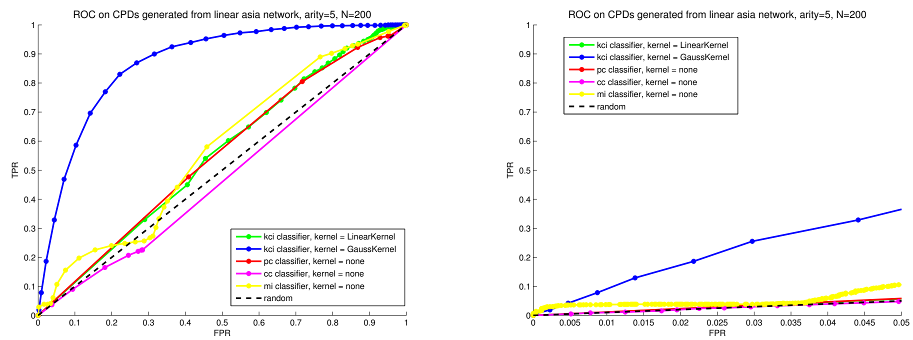

Figure 2: Precision-recall curves for various classifiers. Plots present results for asia network, with nodes having 5 possible classes. Local CPDs have been chosen to express linear relation. **(Left)** Entire precision-recall curve, **(Right)** Low recall fragment of precision-recall curve (XXX).

4.2 Classification of CPDs from Gene Expression

4.3 Synthetic Bayesian networks

4.4 Gene expression data

5 Discussion

References

D. M. Chickering. Learning bayesian networks is np-complete. In *Learning from data*, pages 121–130. Springer, 1996.

L. M. De Campos. A scoring function for learning bayesian networks based on mutual information and conditional independence tests. *The Journal of Machine Learning Research*, 7:2149–2187, 2006.

E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 190–195, 1979.

T. Jaakkola, D. Sontag, A. Globerson, and M. Meila. Learning bayesian network structure using lp relaxations. In *International Conference on Artificial Intelligence and Statistics*, pages 358–365, 2010.

D. Margaritis. *Learning Bayesian network model structure from data*. PhD thesis, University of Pittsburgh, 2003.

R. Opgen-Rhein and K. Strimmer. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology*, 1(1): 37, 2007.

J. Schäfer and K. Strimmer. An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764, 2005.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.