Robust Inference of Structural Independencies from Finite Data

Séverine Affeldt, Hervé Isambert
Institut Curie, Research Center, UMR168, 26, rue d'Ulm, 75005 Paris, France
severine.affeldt@curie.fr, herve.isambert@curie.fr

ABSTRACT

The reconstruction of causal graphical models through constraint-based inference approaches, such as the PC algorithm, is known to be consistent provided that a correct list of conditional independencies is available. Yet, in practice, conditional independencies need to be ascertained from statistical tests on the available observational data, and are not robust to sampling noise from finite datasets. We propose a more robust approach, which uncovers the most likely indirect paths underlying structural independencies, based on the sign and amplitude of conditional 3-point information terms. The resulting `3off2` *algorithm iteratively "takes off" the largest positive conditional 3-point information from the 2-point (mutual) information between each pair of nodes. Then, conditional independencies are derived by progressively collecting the most significant indirect contributions to all pairwise mutual information. Identifying structural independencies within such a maximum likelihood framework is found to be more robust to sampling noise from finite observational data on benchmark networks.*

BACKGROUND

Two types of causal graph inference methods have been developed and applied to a variety of experimental datasets. Bayesian inference approaches have the advantage of allowing for quantitative comparisons between alternative networks using a score. However, as exact bayesian methods are limited to small causal graphs due to the super-exponential space of possible directed graphs to sample, they generally require heuristic search strategies such as hill-climbing algorithms[2].

By contrast, causal inference algorithms based on the identification of structural constraints run in polynomial time on sparse graphs. These constraint-based inference approaches, such as the PC algorithm[10], do not score and compare alternative causal graphs. Instead, they aim at ascertaining conditional independencies between variables to directly infer the markov equivalent class of all causal graphs compatible with the data. Yet, this is usually done in arbitrary order of the considered variables, which is prone to spurious conditional independencies, and is not robust to sampling noise in finite datasets.

NOVEL HYBRID INFERENCE METHOD

We have developed an information-theoretic approach that combines constraint-based and bayesian frameworks to reliably learn graphical models despite of inherent sampling noise in finite datasets. In a nutshell, it ascertains structural independencies in causal graphs (*i.e.*, $I(x;y|\{u_i\}) \sim 0$ implying no x-y link in the underlying network) based on a bayesian ranking of their most contributing nodes, $\{u_i\}$.

In practice, to decide whether a node z should be included in the list of (already identified) contributors, $\{u_i\}$, we rely on the use of (conditional) 3-point information terms, $I(x;y;z|\{u_i\}) = I(x;y|\{u_i\}) - I(x;y|z,\{u_i\})$, which can be positive or negative[7], unlike (conditional) 2-point information terms, $I(x;y|\{u_i\})$, that are always positive. More specifically, it can be shown[1] that the (conditional) 3-point information, $I(x;y;z|\{u_i\})$, is related to the likelihood ratio, $\mathcal{L}_v(xy;z|\{u_i\})/\mathcal{L}_{nv}(xy;z|\{u_i\})$, that xyz form a v-structure ($v \equiv x \rightarrow z \leftarrow y$) *versus* a non v-structure ($nv \equiv x \leftarrow z \leftarrow y$ or $x \leftarrow z \rightarrow y$), where $\{u_i\}$ are already identified contributing nodes of the xy correlations,

$$\mathcal{L}_v(xy;z|\{u_i\})/\mathcal{L}_{nv}(xy;z|\{u_i\}) \quad = \quad e^{-NI(x;y;z|\{u_i\})}$$

Hence, significantly negative (conditional) 3-point information, $I(x;y;z|\{u_i\}) \ll -1/N$, implies that a v-structure is more likely than a non v-structure given the N observed data points. Conversely, significantly positive (conditional) 3-point information, $I(x;y;z|\{u_i\}) \gg 1/N$, implies that a non v-structure model is more likely and that z should be included in the set of nodes, $\{u_i\}$, contributing to the x-y correlations, *i.e.*, $\{u_i\} \leftarrow \{u_i\} + z$, and possibly to the structural independency, *i.e.*, $I(x;y|\{u_i\}) \sim 0$.

By contrast, classical constraint-based approaches[10] assess structural independencies in arbitrary order of the contributing variables, $\{u_i\}$, rendering them prone to spurious conditional independencies. Instead, our novel hybrid approach, `3off2`, progressively uncovers the best supported conditional independencies, by iteratively "taking off" the most significant indirect contributions, the largest positive contributions of conditional 3-point information, $I(x;y;u_k|\{u_i\}_{k-1})$, from every 2-point (mutual) information, $I(x;y)$, of the causal graph, as,

$$\begin{aligned} I(x;y|\{u_i\}_n) \quad &= \quad I(x;y) - I(x;y;u_1) - I(x;y;u_2|u_1) \\ &\quad - \ldots - I(x;y;u_n|\{u_i\}_{n-1}) \end{aligned}$$

Identifying such structural independencies within a maximum likelihood framework proved to be much more robust to sampling noise than classical inference methods and led to excellent results on benchmark networks for large datasets, compared to both Bayesian and constraint-based (PC) methods as well as other information-theoretic approaches, such as Aracne[6].

COMPARISON TO STATE-OF-ART METHODS

We have tested the `3off2` structural inference approach to reconstruct benchmark graphical models containing 20 to 70 nodes and report here the results on ALARM and INSURANCE datasets (see bnlearn database[9] `www.bnlearn.com/`), Fig. 1-2. The undirected skeletons learned by `3off2` have been evaluated against other methods in terms of Precision (or positive predictive value), $Prec = TP/(TP + FP)$, Recall or Sensitivity (or true positive rate), $Rec = TP/(TP + FN)$, as well as F-score(β) for increasing sample size $N = 10$ to $50,000$ data points, where F-score(β)= $(1 + \beta^2)Prec \times Rec(\beta^2 Prec + Rec)$ are evaluated for $\beta = 1/2$ (favoring Precision) and $\beta = 1$ (treating Precision and Recall on the same footing).

Alternative inference methods used are the PC algorithm[10] (significance level $\alpha = 0.01$) implemented in the `pcalg` package[5], and Bayesian inference using the hill-climbing heuristics (with 20 random restarts) and a Bayesian Dirichlet equivalence score[4] implemented in the `bnlearn` package[9]. In addition, we also compare `3off2` to Aracne[6], an information-based inference approach, which iteratively prunes links with the weakest mutual information based on the Data Processing Inequality. We have used the Aracne implementation of the `minet` R package[8], setting the threshold parameter for the minimum difference in mutual information to $\epsilon = 0$. For each sample size, `3off2`, Aracne and PC have been tested on 50 replicates and the Bayesian inference on 20 replicates. Figures 1 and 2 give the average results over these multiple replicates.

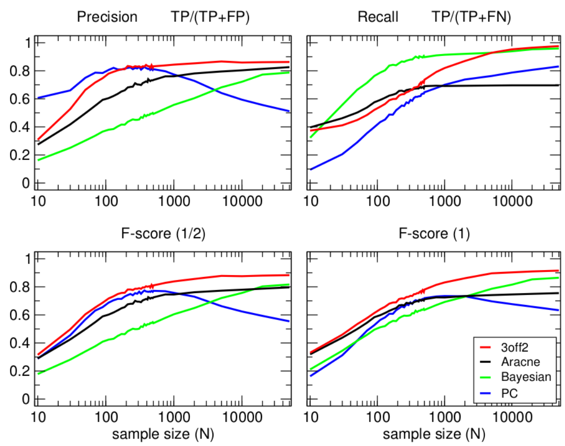

Figure 1. ALARM network (37 **nodes,** 46 **links,** 509 **parameters, Average degree** 2.49**, Maximum in-degree** 4)

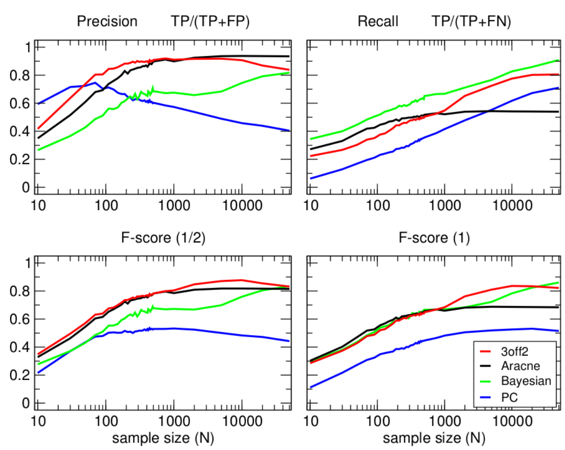

Figure 2. INSURANCE network (27 **nodes,** 52 **links,** 984 **parameters, Average degree** 3.85**, Maximum in-degree** 3)

We have found that the `3off2` approach reaches very good precision levels at smaller sample sizes than Aracne or bayesian methods and, unlike PC, keeps one of the highest Precision score up to large sample sizes $(50,000)$. By contrast, the Recall of `3off2` is lower than with Aracne or bayesian methods at small sample size, although it becomes comparable or better than Aracne and Bayesian inference methods at large sample size. Overall, taking into account Precision and Recall simultaneously results most often in higher `3off2` F-scores as compared to other methods. This is true in particular, at large sample size for F-score($\beta = 1$) and at small sample size for F-score($\beta = 1/2$), as expected for a properly balanced approach favoring Precision over Recall at small sample size when there is usually not enough available data to recover all causal edges.

DISCUSSION, PERSPECTIVE

The `3off2` algorithm exploits the best of constraint-based and Bayesian inference methods to improve the identification of structural constraints of causal graphs from finite datasets. In particular, our approach is expected to run in polynomial time on sparse causal network, like constraint-based algorithms. Besides, the use of local bayesian scores to uncover the best contributing nodes at each iteration enables to reliably identify conditional independencies without cascading accumulation of errors. Among the hybrid methods combining contraint-based and Bayesian approaches, Claassen *et al.*[3] have recently proposed to use bayesian scores to directly assess the reliability of conditional independencies by *summing* the likelihoods over compatible graphs. By contrast, `3off2` circumvents the need to score conditional independencies over a potentially intractable number of graphs by using likelihood ratios while uncovering progressively the best supported conditional independencies.

REFERENCES

1. Affeldt, S., and Isambert, H. Robust inference of conditional independencies from finite data. *submitted* (2014).
2. Chickering, D., Geiger, D., and Heckerman, D. Learning Bayesian networks: Search methods and experimental results. In *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics* (January 1995), 112–128.
3. Claassen, T., and Heskes, T. A bayesian approach to constraint based causal inference. In *In Proc. of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)* (2012), 207–216.
4. Heckerman, D., and Chickering, D. M. Learning bayesian networks: The combination of knowledge and statistical data. In *Machine Learning* (1995), 20–197.
5. Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., and Bühlmann, P. Causal inference using graphical models with the r package pcalg. *Journal of Statistical Software 47*, 11 (2012), 1–26.
6. Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., and Califano, A. Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics 7*, S-1 (2006).
7. McGill, W. J. Multivariate information transmission. (1954). 93–111.
8. Meyer, P. E., Lafitte, F., and Bontempi, G. *minet*: A r/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics 9* (2008).
9. Scutari, M. Learning bayesian networks with the bnlearn r package. *Journal of Statistical Software 35*, 3 (2010), 1–22.
10. Spirtes, P., and Glymour, C. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review 9* (1991), 62–72.