Computational Statistics and Data Analysis 87 (2015) 57–72

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Faithfulness and learning hypergraphs from discrete distributions

Anna Klimova [a,*], Caroline Uhler [a], Tamás Rudas [b]

[a] Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria

[b] Eötvös Loránd University, Pázmány Péter setany 1/A, H-1117 Budapest, Hungary

ARTICLE INFO

Article history:
Received 1 May 2014
Received in revised form 27 January 2015
Accepted 27 January 2015
Available online 3 February 2015

Keywords:
Contingency tables
Directed acyclic graphs
Hierarchical log-linear models
Hypergraphs
(Strong-)faithfulness

ABSTRACT

The concepts of faithfulness and strong-faithfulness are important for statistical learning of graphical models. Graphs are not sufficient for describing the association structure of a discrete distribution. Hypergraphs representing hierarchical log-linear models are considered instead, and the concept of parametric (strong-)faithfulness with respect to a hypergraph is introduced. The strength of association in a discrete distribution can be quantified with various measures, leading to different concepts of strong-faithfulness. It is proven that strong-faithfulness defined in terms of interaction parameters ensures the existence of uniformly consistent parameter estimators and enables building uniformly consistent procedures for a hypergraph search. Lower and upper bounds for the proportions of distributions that do not satisfy strong-faithfulness are computed for different parameterizations and measures of association.

1. Introduction

A graphical model is a set of probability distributions whose association structure can be identified with a graph. Given a graph, the Markov property entails a set of conditional independence relations that are fulfilled by distributions in the model. Distributions in the model that obey no further conditional independence relations are called *faithful to the graph*. For each undirected graphical model, as well as for each directed acyclic graph (DAG) model, there is a distribution that is faithful to the graph (cf. Spirtes et al., 2001). Moreover, the Lebesgue measure of the set of parameters corresponding to distributions that are unfaithful to a graphical model is zero; this result was proven by Spirtes et al. (2001) for the case of multivariate normal distributions, by Meek (1995) for discrete distributions on multi-way contingency tables, and by Peña et al. (2009) for arbitrary sample spaces and dominating measures. It is also well-known, that a DAG model may include distributions that are unfaithful to it but are not Markov to any nested DAG. This kind of unfaithfulness may occur due to path cancellation and can arise both in the discrete and in the multivariate normal settings (cf. Zhang and Spirtes, 2008; Uhler and Raskutti, 2013).

In the discrete case, the non-existence of a graph to which a distribution is faithful is related to the presence of higher than first order interactions in this distribution. Graph learning algorithms (cf. Spirtes et al., 2001), which do not recognize the presence of higher order interactions, may produce a graph which does not reveal the true association structure (cf. Studený, 2005). In order to avoid such errors, graph learning algorithms usually assume the existence of a DAG to which the distribution is faithful. Since the Lebesgue measure of the set of parameters corresponding to distributions that are

* Correspondence to: IST Austria, Am Campus 1, A-3400 Klosterneuburg, Austria. Tel.: +43 0 676 716 2574; fax: +43 0 2243 9000 2000.
E-mail address: aklimova@ist.ac.at (A. Klimova).

http://dx.doi.org/10.1016/j.csda.2015.01.017

unfaithful to the underlying graph is zero, the faithfulness assumption is not considered to be restrictive in the context of graphical search. While graph search procedures assuming faithfulness are pointwise consistent, they are not uniformly consistent and thus cannot simultaneously control Type I and Type II errors with a finite sample size (Robins et al., 2003). To ensure existence of a uniformly consistent learning procedure, strong-faithfulness of a distribution to the underlying DAG is needed (Zhang and Spirtes, 2003). Uhler et al. (2013) analyzed the Gaussian setting and showed that the strong-faithfulness assumption may, in fact, be very restrictive and the corresponding proportions of distributions which do not satisfy strong-faithfulness may become very large as the number of nodes grows.

The concepts of faithfulness and strong-faithfulness were originally introduced in the causal search framework, where they are linked to identifiability of causal effects. However, as we show in this paper using the discrete setting, these concepts are also important for identifiability of more general parameters of association. In Section 2, we define the concept of a model class being closed under a faithfulness relation: for each positive distribution, there exists a model in such a class to which it is faithful. By giving examples of distributions that are not faithful to any directed or undirected graphical model, we show that these model classes are not closed under the faithfulness relation based on the corresponding Markov property. Further, we introduce the concept of parametric faithfulness of a distribution to a hypergraph (instead of a graph). This concept seems more adequate for categorical data, where hypergraphs can be used to represent hierarchical log-linear models. Indeed, we show that the class of models associated with hypergraphs is closed under a parametric faithfulness relation.

In Section 3, we describe two major difficulties with the concept of strong-faithfulness in the discrete case. First, in contrast to role of correlations in the multivariate normal case, there is no single standard measure of the strength of association in a joint distribution. Therefore, depending on the measure of association, different variants of strong-faithfulness may be considered. Second, the proportion of strong-faithful distributions depends on the parameterization used and can only be computed if the parameter space has finite volume. We explore the consequences of different parameterizations and measures of association for the case of the 2×2 contingency table. We define parametric strong-faithfulness with respect to a hypergraph under a parameterization based on the log-linear interaction parameters. Assuming strong-faithfulness, we show that the maximum likelihood estimators of the interaction parameters associated with the hyperedges are uniformly consistent. As a result, we give a set of conditions under which Type I and Type II errors can be controlled with a finite sample size. We also discuss the uniform consistency of model selection procedures for a hypergraph search, for example, using the approaches described by Edwards (2000, 2012).

In Section 4, we estimate the proportion of distributions that do not satisfy the parametric strong-faithfulness assumption with respect to a given hypergraph. We give an exact formulation of these proportions, under a parameterization based on conditional probabilities, for hypergraphs whose hyperedges form a decomposable set. The association structure of such distributions may be discovered incorrectly during a hypergraph learning procedure. Finally, we define the concept of projected strong-faithfulness, which applies to distributions that do not belong to the hypergraph, and estimate the proportions of projected strong-faithful distributions for several hypergraphs over the $2 \times 2 \times 2$ contingency table. In Section 5, we conclude the paper with a brief discussion of our results and their implications.

2. Graphical and parametric faithfulness

We first review the concept of faithfulness with respect to a graph. We then introduce parametric faithfulness with respect to a hypergraph and show that this is a more relevant concept for categorical data.

2.1. Faithfulness with respect to a graph

Let $\mathcal{V}_1, \dots, \mathcal{V}_K$ be random variables taking values in $\mathcal{I} = \mathcal{I}_1 \times \cdots \times \mathcal{I}_K$, a Cartesian product of finite sets. $\mathcal{I}$ describes a K-way contingency table and a vector $\boldsymbol{i} = (i_1, \dots, i_K) \in \mathcal{I}$ forms a cell. A subset $M \subseteq \{1, \dots, K\}$ specifies a marginal of the joint distribution of $\mathcal{V}_1, \dots, \mathcal{V}_K$, and $M = \emptyset$ is the empty marginal. For $M = (k_1, \dots, k_t)$, the set $\mathcal{I}_M = \mathcal{I}_{k_1} \times \cdots \times \mathcal{I}_{k_t}$ is a *marginal table*, and the canonical projection $\boldsymbol{i}_M$ of the cell $\boldsymbol{i}$ onto the set $\mathcal{I}_M$ is a *marginal cell*. We parameterize the population distribution by cell probabilities $\boldsymbol{p} = (p_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathcal{I}}$, where $p_{\boldsymbol{i}} \in (0, 1)$ and $\sum_{\boldsymbol{i} \in \mathcal{I}} p_{\boldsymbol{i}} = 1$, and denote by $\mathcal{P}$ the set of all distributions on $\mathcal{I}$. A subset, $\mathcal{M}$, of $\mathcal{P}$ is called a *model*. For simplicity of exposition, we assume that $\mathcal{V}_1, \dots, \mathcal{V}_K$ are binary, $\mathcal{I}$ is treated as a sequence of cells ordered lexicographically, and a distribution $P \in \mathcal{P}$ is addressed by its parameter, $\boldsymbol{p}$.

A *graphical model* is a set of probability distributions, whose association structure can be identified with a graph with vertices $V = \{1, \dots, K\}$, where each vertex i is associated with a random variable $\mathcal{V}_i$. In the following, we will identify each vertex with its associated random variable. The absence of an edge between two vertices means that the corresponding random variables satisfy some (conditional) independence relation. A detailed description of graphical models for discrete as well as for multivariate normal distributions can be found in Edwards (2000), among others. In the sequel, we only consider undirected graphical models and DAG models.

A graphical model identified with an undirected graph (also called a *graphical log-linear model* in the discrete setting) is a set of probability distributions on V that satisfy the *local undirected Markov property*: Every node is conditionally independent of its non-neighbors given its neighbors. In the discrete case, such models are a sublcass of hierarchical log-linear models. A graphical model identified with a directed acyclic graph, a DAG model, is a set of probability distributions on V that satisfy the *directed Markov property*: Every node is conditionally independent of its non-descendants given its parents. A distribution that satisfies the Markov property with respect to a graph is called *Markov* to it.