Information Sciences 411 (2017) 84–97

Contents lists available at ScienceDirect

Information Sciences

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

Sensitivity analysis in multilinear probabilistic models

Manuele Leonelli[a,*], Christiane Görgen[b], Jim Q. Smith[b]

[a] School of Mathematics and Statistics, University of Glasgow, Glasgow, UK
[b] Department of Statistics, The University of Warwick, Coventry, UK

ARTICLE INFO

Article history:
Received 4 July 2016
Revised 8 March 2017
Accepted 7 May 2017
Available online 10 May 2017

Keywords:
Bayesian networks
CD distance
Interpolating polynomial
Sensitivity analysis
ϕ-divergences

ABSTRACT

Sensitivity methods for the analysis of the outputs of discrete Bayesian networks have been extensively studied and implemented in different software packages. These methods usually focus on the study of sensitivity functions and on the impact of a parameter change to the Chan–Darwiche distance. Although not fully recognized, the majority of these results rely heavily on the multilinear structure of atomic probabilities in terms of the conditional probability parameters associated with this type of network. By defining a statistical model through the polynomial expression of its associated defining conditional probabilities, we develop here a unifying approach to sensitivity methods applicable to a large suite of models including extensions of Bayesian networks, for instance context-specific ones. Our algebraic approach enables us to prove that for models whose defining polynomial is multilinear both the Chan–Darwiche distance and any divergence in the family of ϕ-divergences are minimized for a certain class of multi-parameter contemporaneous variations when parameters are proportionally covaried.

1. Introduction

Many discrete statistical problems in a variety of domains are nowadays often modeled using *Bayesian networks* (BNs) [32]. There are now thousands of practical applications of these models [4,24,26], which have spawned many useful technical developments: including a variety of fast exact, approximate and symbolic propagation algorithms for the computation of probabilities that exploit the underlying graph structure [14,16,17]. Some of these advances have been hard-wired into software [6,27] which has further increased the applicability and success of these methods.

However, BN modeling would not have experienced such a widespread application without tailored methodologies of *model validation*, i.e. checking that a model produces outputs that are in line with current understanding, following a defensible and expected mechanism [19,34]. Such techniques are now well established for BN models [11,27,34]. These are especially fundamental for expert elicited models, where both the probabilities and the covariance structure are defined from the suggestions of domains experts, following knowledge engineering protocols tailored to the BN's building process [30,35]. We can broadly break down the validation process into two steps: the first concerns the auditing of the underlying graphical structure; the second, assuming the graph represents a user's beliefs, checks the impact of the numerical elicited probabilities within this parametric family on outputs of interest. The focus of this paper lies in this second validation phase, usually called a *sensitivity analysis*.

* Corresponding author at:
E-mail addresses: manuele.leonelli@gmail.com, manuele.leonelli@glasgow.ac.uk (M. Leonelli).

http://dx.doi.org/10.1016/j.ins.2017.05.010
0020-0255/

The most common investigation is the so-called *one-way* sensitivity analysis, where the impacts of changes made to a single probability parameter are studied. Analyses where more than one parameter at a time are varied are usually referred to as *multi-way*. In both cases a complete sensitivity analysis for discrete BNs often involves the study of *Chan–Darwiche (CD) distances* [6–8] and *sensitivity functions* [13,40]. The CD distance is used to quantify global changes. It measures how the overall distribution behaves when one (or more) parameter is varied. A significant proportion of research has focused on identifying parameter changes such that the original and the 'varied' BN distributions are close in CD distance [8,37]. This is minimized when, after a single arbitrary parameter change, other covarying parameters, e.g. those from the same conditional distribution, have the same proportion of the residual probability mass as they originally had. Sensitivity functions, on the other hand, model local changes with respect to an output of interest. These describe how that output probability varies as one (or potentially more) parameter is allowed to be changed. Although both these concepts can be applied to generic Bayesian analyses, they have been discussed and applied almost exclusively within the BN literature (see [9,10,36] for some exceptions). This is because the computations of both CD distances and sensitivity functions are particularly straightforward for BN models.

In this paper we introduce a unifying comprehensive framework for certain multi-way analyses, usually called in the context of BNs *single full conditional probability table (CPT) analyses* - where one parameter from each CPT of one vertex of a BN given each configuration of its parents is varied. Using the notion of an interpolating polynomial [33] we are able to describe a large variety of models based on their polynomial form. Then, given this algebraic characterization, we demonstrate that one-way sensitivity methods defined for BNs can be generalized to single full CPT analyses for any model whose interpolating polynomial is multilinear, for example context-specific BNs [3] and stratified chain event graphs [12,39]. Because of both the lack of theoretical results justifying their use and the increase in computational complexity, multi-way methods have not been extensively discussed in the literature: see [2,7,21] for some exceptions. This paper aims at providing a comprehensive theoretical toolbox to start applying such analyses in practice.

Importantly, our polynomial approach enables us to prove that single full CPT analyses in any multilinear model are optimal under proportional covariation in the sense that the CD distance between the original and the varied distributions is minimized. The optimality of this covariation method has been an open problem in the sensitivity analysis literature for quite some time [7,37]. However, we are able to provide further theoretical justifications for the use of proportional covariation in single full CPT analyses. We demonstrate below that for any multilinear model this scheme minimizes not only the CD distance, but also any divergence in the family of ϕ-divergences [1,15]. The class of ϕ-divergences include a very large number of divergences and distances (see e.g. [31] for a review), including the famous Kullback–Leibler (KL) divergence [28]. The application of KL distances in sensitivity analyses of BNs has been almost exclusively restricted to the case when the underlying distribution is assumed Gaussian [20,21], because in discrete BNs the computation of such a divergence requires more computational power than for CD distances. We demonstrate below that this additional complexity is a feature shared by any divergence in the family of ϕ-divergences.

The paper is structured as follows. In Section 2 we define interpolating polynomials and demonstrate that commonly used models entertain a polynomial representation. In Section 3 we review a variety of divergence measures. Section 4 presents a variety of results for single full CPT sensitivity analyses in multilinear models, namely the derivation of sensitivity functions and the proof of optimality of proportional covariation. We conclude with a discussion.

2. Multilinear and polynomial parametric models

In this section we first provide a generic definition of a parametric statistical model together with the notion of interpolating polynomial. We then discuss parametric models whose interpolating polynomial is multilinear.

2.1. Parametric models and interpolating polynomials

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be a random vector with an associated discrete and finite sample space $\mathbb{Y}$, with $\#\mathbb{Y} = q$. Although our methods straightforwardly applies when the entries of $\boldsymbol{Y}$ are random vectors, for ease of notation, we henceforth assume its elements are univariate.

Definition 1. Denote by $\mathbb{P}_{\boldsymbol{\theta}} = \{p_{\boldsymbol{\theta}}(\boldsymbol{y}) \mid \boldsymbol{y} \in \mathbb{Y}\}$ the values of a probability mass function $p_{\boldsymbol{\theta}} : \mathbb{Y} \to [0, 1]$ which depends on a choice of parameters $\boldsymbol{\theta} \in \mathbb{R}^k$. The entries of $\mathbb{P}_{\boldsymbol{\theta}}$ are called *atomic probabilities* and the elements $\boldsymbol{y} \in \mathbb{Y}$ *atoms*.

Definition 2. A discrete *parametric statistical model* on $q \in \mathbb{N}$ atoms is a subset $\mathbb{P}_\Psi \subseteq \Delta_{q-1}$ of the $q-1$ dimensional probability simplex, where

$$\Psi \; : \quad \mathbb{R}^k \to \mathbb{P}_\Psi, \; \boldsymbol{\theta} \mapsto \mathbb{P}_{\boldsymbol{\theta}}, \tag{1}$$

is a bijective map identifying a particular choice of parameters $\boldsymbol{\theta} \in \mathbb{R}^k$ with one vector of atomic probabilities. The map Ψ is called a *parameterisation* of the model.

The above definition is often encountered in the field of *algebraic statistics*, where properties of statistical models are studied using techniques from algebraic geometry and commutative computer algebra, among others [18,38]. We next follow [22] in extending some standard terminology.