International Journal of Approximate Reasoning 55 (2014) 1022–1042

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Co-variation for sensitivity analysis in Bayesian networks: Properties, consequences and alternatives

Silja Renooij *

Department of Information and Computing Sciences, Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

ARTICLE INFO

Article history:
Available online 16 July 2013

Keywords:
Sensitivity analysis
Bayesian networks
Co-variation
Order-preserving
CD-distance
CD-lowerbound

ABSTRACT

Upon varying parameters in a sensitivity analysis of a Bayesian network, the standard approach is to co-vary the parameters from the same conditional distribution such that their proportions remain the same. Alternative co-variation schemes are, however, possible. In this paper we investigate the properties of the standard proportional co-variation and introduce two alternative schemes: uniform and order-preserving co-variation. We theoretically investigate the effects of using alternative co-variation schemes on the so-called sensitivity function, and conclude that its general form remains the same under any linear co-variation scheme. In addition, we generalise the CD-distance for bounding global belief change to explicitly include the co-variation scheme under consideration. We prove a tight lower bound on this distance for parameter changes in single conditional probability tables.

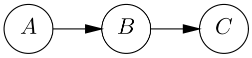
1. Introduction

Sensitivity analysis is a general technique for studying the effects of parameter changes on the output of a mathematical model. In the context of Bayesian networks the output of interest can be any probability computed from the network; the parameters to which changes are applied consist of one or more probabilities from the network's conditional probability tables. The results of a sensitivity analysis can be captured in detail by means of a *sensitivity function*, describing an output probability of interest as a function of one or more parameter probabilities [1]. More global effects of parameter changes can be described by the *CD-distance* [2]. The CD-distance is a measure for bounding probabilistic belief change and complements the sensitivity function by giving insight in the effect of parameter changes on the global joint distribution, rather than on a specific (posterior) output probability of interest.

Upon varying a probability from a (conditional) distribution, the remaining probabilities from the same distribution need to be co-varied. The *proportional scheme* has been adopted as the standard scheme for co-variation in Bayesian networks, and various algorithms and properties associated with sensitivity analysis build upon this scheme [1–5]. For example, the known standard form of the sensitivity function is based on proportional co-variation [1]. The proportional co-variation scheme, however, is one of numerous alternatives for co-varying parameters from the same distribution. In this paper we investigate the properties of the proportional co-variation scheme and argue that the mere fact that it is the standard co-variation scheme used, does not imply that there are no situations in which alternative schemes may be more suitable. For example, we may want to prevent certain parameters from co-varying, or preserve some relation — such as the order — between parameters. We therefore introduce some alternative schemes, study their properties and compare sensitivity functions established under the different schemes.

* Tel.: +31 30 2539266; fax: +31 30 2532804.
E-mail address: S.Renooij@uu.nl.

0888-613X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2013.07.004



Θ_A	a_1	a_2
	0.9	0.1

$\Theta_{B\|A}$	b_1	b_2	b_3	b_4
a_1	0.2	0.2	0.1	0.5
a_2	0.1	0.3	0.4	0.2

$\Theta_{C\|B}$	c_1	c_2
b_1	0.01	0.99
b_2	0.5	0.5
b_3	0.9	0.1
b_4	0.2	0.8

Fig. 1. Example Bayesian network with three nodes and their CPTs.

A known property of the proportional co-variation scheme is that it is optimal in the context of varying a single parameter, in the sense that it minimises the CD-distance between the original and the new distribution [2]. It is as of yet unknown, however, if the proportional scheme is also optimal when multiple, independent parameters are varied. Moreover, the choice of the CD-distance as the measure to optimise in the context of sensitivity analysis is rather arbitrary: we could also be interested in minimising the well-known KL-divergence [6], or some other measure. Parameter changes that minimise KL-divergence do not necessarily minimise CD-distance, or vice versa [2]. In addition, to get a 'worst case' impression, we may want to perform our analyses in the context of large disturbances to the network, rather than minimal ones. In this paper we will therefore investigate exactly how both the sensitivity function and the CD-distance depend on the co-variation scheme used. We show that the general form of the sensitivity function is maintained as long as the co-variation scheme is linear in the parameter(s) varied. In addition, we generalise the CD-distance to arbitrary co-variation schemes, and prove that a previously suggested approximation of this distance is in fact a lower bound.

This paper is organised as follows. Section 2 provides preliminaries on Bayesian networks and sensitivity analysis. Section 3 discusses proportional co-variation, introduces alternatives and studies properties of the various schemes. In Section 4, we generalise the sensitivity function to arbitrary co-variation schemes; likewise, Section 5 generalises the CD-distance. The paper ends with conclusions and directions for future research in Section 6.

2. Preliminaries

In this section we briefly review Bayesian networks and sensitivity analysis in such networks.

2.1. Bayesian networks

A Bayesian network compactly represents a joint probability distribution Pr over a set of discrete stochastic variables $\mathbf{V}$ [7,8]. It combines an acyclic directed graph G, that captures the variables and their dependencies as nodes and arcs respectively, with conditional probability distributions $\Theta_{V_i|\mathbf{U}}$ for each variable V_i and its (possibly empty set of) parents $\mathbf{U} = \pi(V_i)$ in the graph. An example Bayesian network is shown in Fig. 1.

A Bayesian network uniquely defines the joint distribution

$$\Pr(\mathbf{V}) = \prod_i \Theta_{V_i|\mathbf{U}}$$

that respects the probabilistic independences read from the digraph G by means of the d-separation criterion [8]. A Bayesian network, together with its associated inference algorithms, allows for computing any probability of interest over its variables.

In the remainder of this paper, we will refer to $\Theta_{V_i|\mathbf{U}}$ as the conditional probability table (CPT) of variable or node V_i; entries θ of Θ are called parameter probabilities, or parameters for short. Note that each row in a CPT captures a single distribution $\Theta_{V_i|\mathbf{u}}$ over V_i, for a specific combination of values $\mathbf{u}$ for its parents. Variables are denoted by capital letters and their values or instantiations by lower case; bold face is used for sets.

2.2. Sensitivity analysis

Probabilities computed from a Bayesian network are affected by the inevitable inaccuracies in the network's parameters [9]. To investigate the extent of these effects, a sensitivity analysis can be performed [1,3–5,10,11]. In a so-called *n-way* sensitivity analysis of a Bayesian network, $n \geqslant 1$ network parameters are varied simultaneously and the effects on output probabilities of interest are studied.

The most often performed sensitivity analysis is a 1-way analysis, in which a single parameter $\theta_{v_i|\mathbf{u}}$ from a single distribution $\Theta_{V|\mathbf{u}}$ in a single CPT $\Theta_{V|\mathbf{U}}$ is explicitly varied; other parameters $\theta_{v_j|\mathbf{u}}$ from the same conditional distribution are necessarily *co-varied* to ensure that their sum remains 1. In general, an n-way analysis can involve parameters from at most n different CPTs [4,5]. Since such an analysis can quickly become computationally infeasible as n grows, in practice n-way analyses are typically restricted to very small n (2, at most 3), or to n parameters from a single CPT [4]. Moreover, to avoid having to impose additional constraints on the varied parameters, it is often assumed that at most one parameter per distribution is varied explicitly [7].