International Journal of Approximate Reasoning 52 (2011) 805–818

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

# Extended Shenoy–Shafer architecture for inference in hybrid bayesian networks with deterministic conditionals

Prakash P. Shenoy *, James C. West

*University of Kansas School of Business, 1300 Sunnyside Ave, Summerfield Hall, Lawrence, KS 66045-7601, USA*

ARTICLE INFO

*Article history:*
Received 23 June 2010
Received in revised form 10 December 2010
Accepted 14 February 2011
Available online 5 March 2011

*Keywords:*
Hybrid Bayesian networks
Inference in hybrid Bayesian networks
Shenoy–Shafer architecture
Extended Shenoy–Shafer architecture
Dirac delta functions

ABSTRACT

The main goal of this paper is to describe an architecture for solving large general hybrid Bayesian networks (BNs) with deterministic conditionals for continuous variables using local computation. In the presence of deterministic conditionals for continuous variables, we have to deal with the non-existence of the joint density function for the continuous variables. We represent deterministic conditional distributions for continuous variables using Dirac delta functions. Using the properties of Dirac delta functions, we can deal with a large class of deterministic functions. The architecture we develop is an extension of the Shenoy–Shafer architecture for discrete BNs. We extend the definitions of potentials to include conditional probability density functions and deterministic conditionals for continuous variables. We keep track of the units of continuous potentials. Inference in hybrid BNs is then done in the same way as in discrete BNs but by using discrete and continuous potentials and the extended definitions of combination and marginalization. We describe several small examples to illustrate our architecture. In addition, we solve exactly an extended version of the crop problem that includes non-conditional linear Gaussian distributions and non-linear deterministic functions.

Published by Elsevier Inc.

## 1. Introduction

Bayesian networks (BNs) and influence diagrams (IDs) were invented in the mid 1980s (see e.g., [25,10]) to represent and reason with large multivariate discrete probability models and decision problems, respectively. Several efficient algorithms exist to compute exact marginals of posterior distributions for discrete BNs (see e.g., [15,37]) and to solve discrete influence diagrams exactly (see e.g., [24,30,33]).

The state of the art exact algorithm for mixtures of Gaussians hybrid BNs is Lauritzen–Jensen's [16] algorithm implemented with [20] lazy propagation technique. This requires the conditional distributions of continuous variables to be conditional linear Gaussians, and that discrete variables do not have continuous parents. Marginals of multivariate normal distributions can be found easily without the need for integration. The disadvantages are that in the inference process, continuous variables have to be marginalized before discrete ones. In some problems, this restriction can lead to large cliques [18].

If a BN has discrete variables with continuous parents, Murphy [22] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a conditional linear Gaussian. Lerner [17] uses a numerical integration technique called Gaussian quadrature to approximate non-conditional linear Gaussian distributions with conditional linear Gaussians, and this same technique can be used to approximate the product of potentials associated

* Corresponding author.
*E-mail addresses:* pshenoy@ku.edu (P.P. Shenoy), cully@ku.edu (J.C. West).

0888-613X/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ijar.2011.02.005

with a discrete variable and its continuous parents. Murphy's and Lerner's approach is then embedded in Lauritzen–Jensen's [16] algorithm to solve the resulting mixtures of Gaussians BN.

Shenoy [35] proposes approximating non-conditional linear Gaussian distributions by mixtures of Gaussians using a non-linear optimization technique, and using arc reversals to ensure discrete variables do not have continuous parents. The resulting mixture of Gaussians BN is then solved using Lauritzen–Jensen's [16] algorithm.

Moral et al. [21] proposes approximating probability density functions (PDFs) by mixtures of truncated exponentials (MTE), which are easy to integrate in closed form. Since the family of mixtures of truncated exponentials are closed under combination and marginalization, the Shenoy–Shafer [37] algorithm can be used to solve a MTE BN. Cobb and Shenoy [4] and Cobb et al. [6] propose using a non-linear optimization technique for finding mixtures of truncated exponentials approximation for the many commonly used distributions. Cobb and Shenoy [2,3] extend this approach to BNs with linear and non-linear deterministic variables. In the latter case, they approximate non-linear deterministic functions by piecewise linear ones. Rumi and Salmeron [28] describe approximate probability propagation with MTE approximations that have only two exponential terms in each piece. Romero et al. [27] describe learning MTE potentials from data, and Langseth et al. [14] investigate the use of MTE approximations where the coefficients are restricted to integers.

Shenoy and West [39] have proposed mixtures of polynomials, in the same spirit as MTEs, as a solution to the integration problem. Shenoy [36] proposes relaxing the hypercube condition of MOP functions, which enables easy representation of two and three-dimensional CLG conditionals by MOP functions. The family of MOP functions is closed under transformations needed for multi-dimensional linear and quotient deterministic functions.

For Bayesian decision problems, Kenley [12] (see also [32]) describes the representation and solution of Gaussian IDs that include continuous chance variables with conditional linear Gaussian distributions. Poland [26] extends Gaussian IDs to mixture of Gaussians IDs. Thus, continuous chance variables can have any distributions, and these are approximated by mixtures of Gaussians. Cobb and Shenoy [5] extend MTE BNs to MTE IDs for the special case where all decision variables are discrete. Li and Shenoy [19] have proposed an architecture that is an extension of the architecture described in this paper for solving hybrid influence diagrams with deterministic variables.

In this paper, we describe a generalization of the Shenoy–Shafer architecture for discrete BNs so that it applies to hybrid BNs with deterministic conditionals for continuous variables. The functions associated with deterministic conditionals do not have to be linear (as in the CLG case) or even invertible. We use Dirac delta functions to represent such functions. We keep track of the units of continuous potentials. This enables us, e.g., to describe the units of the normalization constant, which are often referred to as "probability" of evidence. Finally, we illustrate our architecture using several small examples, and by solving a modified version of the crop problem initially introduced by Murphy [22].

An outline of the remainder of the paper is as follows. In Section 2, we define Dirac delta functions and describe some of their properties. In Section 3, we describe our architecture for making inferences in hybrid BNs with deterministic variables. This is the main contribution of this paper. In Section 4, we describe four small examples of hybrid BNs with deterministic variables to illustrate our definitions and our architecture. In Section 5, we describe and solve a modification of the crop problem, initially described by Murphy [22], and subsequently modified by a number of authors. Finally, in Section 6, we end with a summary and discussion.

## 2. Dirac delta functions

In this section, we define Dirac delta functions. We use Dirac delta functions to represent deterministic conditionals associated with some continuous variables in BNs. Dirac delta functions are also used to represent observations of continuous variables.

$\delta : \mathbb{R} \to \mathbb{R}^+$ is called a *Dirac delta function* if $\delta(x) = 0$ if $x \neq 0$, and $\int \delta(x)\mathrm{d}x = 1$. Whenever the limits of integration of an integral are not specified, the entire range $(-\infty, \infty)$ is to be understood. The values of $\delta$ are assumed to be in units of density. $\delta$ is not a proper function since the value of the function at 0 does not exist (i.e., is not finite). It can be regarded as a limit of a certain sequence of functions (such as, e.g., the Gaussian density function with mean 0 and variance $\sigma^2$ in the limit as $\sigma \to 0$). However, it can be used as if it were a proper function for practically all our purposes without getting incorrect results. It was first defined by Dirac [7].

As defined above, the value $\delta(0)$ is undefined, i.e., $\infty$, in units of density. We argue that we can *interpret* the value $\delta(0)$ as probability 1. Consider the normal PDF with mean 0 and variance $\sigma^2$. Its moment generating function (MGF) is $M(t) = e^{\sigma^2 t^2/2}$. In the limit as $\sigma \to 0$, $M(t) = 1$. Now, $M(t) = 1$ is the MGF of the distribution $X = 0$ with probability 1. Therefore, we can *interpret* the value $\delta(0)$ (in units of density) as probability 1 at the location $x = 0$.

Some basic properties of the Dirac delta functions that are useful in uncertain reasoning are described in the Appendix. Properties (i)–(iv) are useful in integrating potentials containing Dirac delta functions. Property (v) defines the Heaviside function, which is related to the Dirac delta function. Properties (vi)–(x) are useful in representing deterministic conditionals by Dirac delta functions.

Consider a simple Bayesian network consisting of two continuous variables $X$ and $Y$ with $X$ as a parent of $Y$. Suppose $X$ has PDF $f_X(x)$, and suppose the conditional PDF for $Y$ given $X = x$ is given by $f_{Y|x}(y)$. Then, it follows from probability theory that the marginal for $Y$ can be found by first multiplying the two PDFs to yield the joint PDF of $X$ and $Y$, and then integrating $X$ from the joint. Thus, if $f_Y(y)$ denotes the marginal of $Y$,