International Journal of Approximate Reasoning 53 (2012) 847–866

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

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

Two issues in using mixtures of polynomials for inference in hybrid Bayesian networks

Prakash P. Shenoy

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

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 24 January 2012
Accepted 25 January 2012
Available online 7 February 2012

Keywords:
Inference in hybrid Bayesian networks
Mixtures of polynomials
Conditional linear Gaussian distributions
Lagrange interpolating polynomials
Chebyshev points
Conditional log-normal distributions

ABSTRACT

We discuss two issues in using mixtures of polynomials (MOPs) for inference in hybrid Bayesian networks. MOPs were proposed by Shenoy and West for mitigating the problem of integration in inference in hybrid Bayesian networks. First, in defining MOP for multi-dimensional functions, one requirement is that the pieces where the polynomials are defined are hypercubes. In this paper, we discuss relaxing this condition so that each piece is defined on regions called hyper-rhombuses. This relaxation means that MOPs are closed under transformations required for multi-dimensional linear deterministic conditionals, such as $Z = X + Y$, etc. Also, this relaxation allows us to construct MOP approximations of the probability density functions (PDFs) of the multi-dimensional conditional linear Gaussian distributions using a MOP approximation of the PDF of the univariate standard normal distribution. Second, Shenoy and West suggest using the Taylor series expansion of differentiable functions for finding MOP approximations of PDFs. In this paper, we describe a new method for finding MOP approximations based on Lagrange interpolating polynomials (LIP) with Chebyshev points. We describe how the LIP method can be used to find efficient MOP approximations of PDFs. We illustrate our methods using conditional linear Gaussian PDFs in one, two, and three dimensions, and conditional log-normal PDFs in one and two dimensions. We compare the efficiencies of the hyper-rhombus condition with the hypercube condition. Also, we compare the LIP method with the Taylor series method.

1. Introduction

An hybrid Bayesian network (BN) is a BN with a mix of discrete and continuous random variables. A random variable is said to be *discrete* if its state space is countable, and *continuous* otherwise. Each variable in a BN is associated with a conditional distribution function (conditional, in short) for the variable given its parents. A conditional is said to be *deterministic* if its conditional variance is zero for each state of its parents.

Marginalizing a continuous variable involves integration of the product of all potentials that contain the variable in their domains. Often, these potentials are not integrable in closed form. This is a major problem in making inferences in hybrid BNs. We will call this the *integration* problem.

Literature review. A traditional approximate method for dealing with the integration problem is to discretize all continuous variables. If the number of bins used for discretization is large (to increase the accuracy of the results), the computational effort required to find marginals can be large. A priori, we may not know the regions of the continuous variables where the posterior density lies. Kozlov and Koller [1] have proposed a dynamic discretization technique where one starts with a uniform coarse discretization, and then iteratively refines the discretization based on the location of the probability masses.

E-mail address: pshenoy@ku.edu

0888-613X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2012.01.008

Another approximate method for dealing with the integration method is to use Monte Carlo sampling methods. There are a host of methods including importance sampling (e.g., [2,3]) and Markov chain Monte Carlo (e.g., [4]). The idea is to sample from the posterior distribution. In the presence of deterministic conditionals, convergence can be a problem.

One exact solution to the integration problem proposed by Lauritzen and Jensen [5] is to restrict conditionals of continuous variables to the conditional linear Gaussian (CLG) family, and for discrete variables to not have continuous parents. Such BNs are called *mixture of Gaussians* BNs. In this case, we can avoid the integration problem as marginals of multivariate normal distributions are multivariate normal and no integration needs to be done. However, restricting conditionals to the CLG family can be too restrictive. Also, the requirement that discrete variables not have continuous parents can also be too restrictive. Finally, in finding marginals, all continuous variable have to be marginalized before marginalizing discrete ones, and this restriction can lead to large cliques making inference intractable [6].

If a BN has discrete variables with continuous parents, Murphy [7] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a CLG distribution. Lerner et al. [8] uses a numerical integration technique called Gaussian quadrature to approximate non-CLG distributions with CLG distributions, and this same technique can be used to approximate the product of potentials associated with a discrete variable and its continuous parents. Murphy's and Lerner's approach is then embedded in the Lauritzen-Jensen [5] algorithm to solve the resulting mixtures of Gaussians BN.

Shenoy [9] proposes approximating non-CLG distributions by mixtures of Gaussians using a nonlinear 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 the Lauritzen-Jensen [5] algorithm.

Another solution to the integration problem is to approximate conditional PDFs by a functions called *mixtures of truncated exponentials* (MTEs) [10]. MTE functions are piecewise functions that are defined on regions called hypercubes, and the functions themselves are exponential functions of a linear function of the variables. Such functions are easy to integrate, and the family of MTE functions is closed under multiplication, addition, and integration, three operations that are used in finding marginals using the extended Shenoy-Shafer architecture [11]. Cobb et al. [12] describe MTE approximations of several commonly used one-dimensional PDFs. Moral et al. [13] describe a mixed-tree method for representing an MTE approximation of the 2-dimensional CLG distribution. Parameter learning in MTE networks are discussed in [14,15]. Rumí and Salmerón [16] discuss approximate inference in MTE hybrid BNs that do not contain deterministic conditionals.

Another method that is similar in principle to the MTE method is the mixture of polynomials (MOP) method proposed by Shenoy and West [17]. Instead of using piecewise exponential functions, the MOP method uses piecewise polynomials. Although a detailed comparison of MTE and MOP methods has yet to be done, an advantage of the MOP method is that one can easily find MOP approximations of differentiable PDFs using the Taylor series expansion of the PDF. Shenoy and West [17] describe a MOP approximation of a two-dimensional CLG distribution using the Taylor series method.

Contributions. In both the MTE and the MOP methods, the multi-dimensional piecewise functions are defined on regions called hypercubes. One advantage of this restriction is that such multi-dimensional piecewise functions are easy to integrate. However, the hypercube restriction poses two limitations. First, it is difficult to find an MTE or a MOP approximation of a multi-dimensional conditional PDF for dimensions greater than two. The mixed-tree method proposed by Moral et al. [13] and the Taylor series method proposed by Shenoy and West [17] do not scale up to higher dimensions in practice, i.e., the approximations using these methods have too many pieces or too many terms or have too high a degree for practical use.

The second limitation is that in the presence of multi-dimensional linear deterministic conditionals, the family of MTE and MOP functions are not closed. For example, suppose X has PDF $f_X(x)$ and suppose Y has conditional PDF $f_{Y|x}(y)$, and suppose Z has a deterministic conditional given by the linear function $Z = X + Y$. To find the marginal distribution of Z, we need to combine $f_X(x)$ and $f_{Y|x}(z-x)$ and then integrate x out of the combination. The problem is that even if $f_{Y|x}(y)$ was defined on hypercubes, $f_{Y|x}(z-x)$ is no longer defined on hypercubes. This problem applies equally to the MTE and MOP methods.

In this paper, we suggest replacing the hypercube condition with a more general hyper-rhombus condition. For one-dimensional functions, the two conditions coincide. However, for dimensions two or greater, the hyper-rhombus condition is a generalization of the hypercube condition. The hyper-rhombus condition has three important advantages. First, MOP functions defined on hyper-rhombuses are closed under operations required for multi-dimensional linear deterministic conditionals. Second, it allows us to define MOP approximations of high-dimensional CLG distributions using a MOP approximation of the one-dimensional standard normal PDF. Third, the hyper-rhombus condition allows us to find MOP approximations of multi-dimensional conditional PDFs that have fewer pieces and lower degrees than MOP approximations that are restricted to hypercubes.

Another contribution of this paper is a method for finding MOP approximations of PDFs based on Lagrange interpolating polynomials (LIP) with Chebyshev points. We describe this method, and compare it with the Taylor series method. The LIP method produces MOP approximations that have a better fit than the Taylor series method assuming the same number of pieces and same degree. The LIP method does not require a PDF to be differentiable. For multi-dimensional conditional PDFs, the LIP method with Chebyshev points coupled with the hyper-rhombus condition allows us to find MOP approximations that have fewer pieces and lower degrees than MOP approximations found using Taylor series method.