Project no.: **619209**

Project full title: **Analysis of Massive Data STreams**

Project Acronym: **AMIDST**

Deliverable no.: **D3.1**

Title of the deliverable: **State of the art of inference in hybrid and dynamic models**

Contractual Date of Delivery to the CEC:	**31.10.2014**
Actual Date of Delivery to the CEC:	**31.10.2014**
Organisation name of lead contractor for this deliverable:	**UAL**
Author(s):	**Helge Langseth, Anders L. Madsen, Thomas D. Nielsen, Rafael Rumí, Antonio Salmerón**
Participants(s):	**P01, P02, P03, P04**
Work package contributing to the deliverable:	**WP3**
Nature:	**R**
Version:	**1.0**
Total number of pages:	**26**
Start date of project:	**1st January 2014 Duration: 36 month**

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

The aim of this document is to provide a thorough and up-to-date description of the state of the art on inference in hybrid Bayesian networks. This is intended to serve as a basis for the developments to be carried out in WP3. The document covers inference in static models, MAP inference and inference in dynamic models. The included references have been analysed taking into account the distributional models they utilise and the nature of the provided solutions, either exact or approximate, as well as the potential strengths and drawbacks from the point of view of the objectives of WP3.

Keyword list: hybrid Bayesian networks, inference in static domains, MAP inference, inference in dynamic domains.

Contents

Document history

Version	Date	Author (Unit)	Description
v0.3	8/9 2014	Helge Langseth, Anders L. Madsen, Thomas D. Nielsen, Rafael Rumí, Antonio Salmerón	First draft finished
v0.6	27/10 2014	Helge Langseth, Anders L. Madsen, Thomas D. Nielsen, Rafael Rumí, Antonio Salmerón	Initial draft finished and reviewed by the PSRG
v1.0	30/10 2014	Helge Langseth, Anders L. Madsen, Thomas D. Nielsen, Rafael Rumí, Antonio Salmerón	Final version of document

1 Executive summary

The aim of this document is to provide a thorough and up-to-date description of the state of the art on inference in hybrid Bayesian networks. This is intended to serve as a basis for the developments to be carried out in WP3.

The document has been structured according to the task division in WP3, with separate sections covering inference in static models, MAP inference and inference in dynamic models. The included references have been analysed taking into account the distributional models they utilise and the nature of the provided solutions, either exact or approximate, as well as the potential strengths and drawbacks from the point of view of the objectives of WP3.

2 Introduction

Probabilistic graphical models provide a well-founded and principled approach for performing inference in complex domains endowed with uncertainty. A probabilistic graphical model is a framework consisting of two parts: a qualitative component in the form of a graphical model encoding conditional independence assertions about the domain being modelled as well as a quantitative component consisting of a collection of local probability distributions adhering to the independence properties specified in the graphical model. Collectively, the two components provide a compact representation of the joint probability distribution over the domain being modelled.

Bayesian networks (BNs) [1] are a particular type of probabilistic graphical model that has enjoyed widespread attention in the last two decades. Figure 1 shows a BN representing the joint distribution of variables $X_1, \ldots, X_5$. Attached to each node, there is a conditional probability distribution given its parents in the network, so that the joint distribution factorises as

$$p(X_1, \ldots, X_5) = p(X_1)p(X_2|X_1)p(X_3|X_1)p(X_4|X_2, X_3)p(X_5|X_3).$$

In general, for a BN with n variables $\mathbf{X} = \{X_1, \ldots, X_N\}$, the joint distribution factorises as

$$p(\mathbf{X}) = \prod_{i=1}^{N} p(X_i|Pa(X_i)), \tag{1}$$

where $Pa(X_i)$ denotes the set of parents of X_i in the network. A BN is called *hybrid* if some of its variables are discrete while some others are continuous.

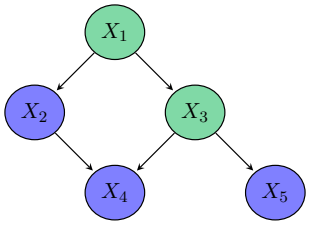

Figure 1: A Bayesian network with two discrete and three continuous variables.

We will use lowercase letters to refer to values or configurations of values, so that x denotes a value of X and $\mathbf{x}$ is a configuration of the variables in $\mathbf{X}$. Given a set of observed variables $\mathbf{X}_E \subset \mathbf{X}$ and a set of variables of interest $\mathbf{X}_I \subset \mathbf{X} \setminus \mathbf{X}_E$, *probabilistic inference* consists of computing the posterior distribution $p(x_i|\mathbf{x}_E)$ for each $i \in I$. If we denote by $\mathbf{X}_C$ and $\mathbf{X}_D$ the set of continuous and discrete variables not in $\{\mathbf{X}_i\} \cup \mathbf{X}_E$,

and by $\mathbf{X}_{C_i}$ and $\mathbf{X}_{D_i}$ the set of continuous and discrete variables not in $\mathbf{X}_E$, the goal of inference can be formulated as computing

$$p(x_i|\mathbf{x}_E) = \frac{p(x_i, \mathbf{x}_E)}{p(\mathbf{x}_E)} = \frac{\sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} \int_{\mathbf{x}_C \in \Omega_{\mathbf{X}_C}} p(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_C}{\sum_{\mathbf{x}_{D_i} \in \Omega_{\mathbf{X}_{D_i}}} \int_{\mathbf{x}_{C_i} \in \Omega_{\mathbf{X}_{C_i}}} p(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_{C_i}}, \tag{2}$$

where $\Omega_{\mathbf{X}}$ stands for the set of possible values of a set of variables $\mathbf{X}$ and $p(\mathbf{x}, \mathbf{x}_E)$ is the joint distribution in the BN instantiated to the observed values $\mathbf{x}_E$.

Inference is often carried out over an auxiliary structure called a *join tree* or *junction tree*, which is an undirected tree constructed from the original BN where the nodes are the cliques of a triangulated graph obtained from the BN. Each conditional distribution is assigned to a node in the join tree and inference can be organised by passing messages throughout the tree.

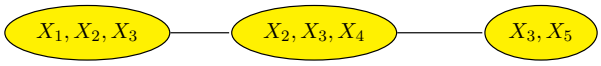

Figure 2: A join tree constructed from the BN in Fig. 1.

The complexity of inference in BNs is motivated by the difficulty in handling the joint distribution $p(\mathbf{x}, \mathbf{x}_E)$ even if it is compactly represented by the factorisation induced by the BN structure. In general, the computational complexity of both exact and approximate inference in BNs is NP-hard [2,3], but many real-world applications can nonetheless be solved efficiently by state-of-the-art inference algorithms [4–7].

Although one cannot hope to achieve polynomial time performance when designing and adapting inference algorithms, we can still aim for polynomial or even linear complexity that will allow for the evaluation of more complex domains. One approach for achieving this is to exploit parallel architectures, where several processors/cores can be used to simultaneously solve different parts of the problem and afterwards obtain a global solution by combining the solutions to the parts [8–14]. It is also possible to adopt simplified models either by following modelling tricks [15] or by using fixed structures, especially if the model is aimed at specific problems like classification [16]. Complexity reduction can also be attained by simplifying the structure of the network [17] or even by using compact representations of the probabilistic potentials in the network [18].

A particularly complex kind of inference in BNs is the so-called *maximum a posteriori (MAP)* problem. For a set of target variables $\mathbf{X}_I \subset \mathbf{X} \setminus \mathbf{X}_E$, the goal of MAP inference is to compute

$$\mathbf{x}_I^* = \arg \max_{\mathbf{x}_I \in \Omega_{\mathbf{X}_I}} p(\mathbf{x}_I | \mathbf{X}_E = \mathbf{x}_E), \tag{3}$$

where $p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$ is obtained by first marginalising out from the joint distribution $p(\mathbf{x})$ the variables not in $\mathbf{X}_I$ and not in $\mathbf{X}_E$, similarly to Eq. (2). A related problem is *MPE* that stands for finding the *most probable explanation* to an observation $\mathbf{X}_E = \mathbf{x}_E$. It is a particular case of MAP, where $\mathbf{X}_I = \mathbf{X} \setminus \mathbf{X}_E$. Both MAP and MPE belong to the class of problems known as *abductive inference* [19].

The increased complexity of MAP inference with respect to standard inference is due to the fact that evaluating Eq. 3 involves maximisation and marginalisation (either by summation or integration), and marginalisation has to be performed prior to maximisation. Such restrictions on the order of the calculations can force inference algorithms to work with potentials of larger size than during standard inference.

MAP inference is known to be NP-hard even for simple network structures like poly trees and naïve Bayes [20, 21]. This makes approximate algorithms even more necessary for handling practical situations. State-of-the-art approximate algorithms are based on simplifying the problem structure and refining the solution afterwards [22].

3 Inference in static hybrid Bayesian networks

Traditionally, Bayesian networks have been defined for discrete domains, where the entities of interest are modelled by discrete variables which ensures that belief updating can be performed efficiently and in closed form. However, this requirement also imposes severe restrictions as many domains contain entities that are more appropriately modelled by variables with continuous state spaces; an example is distance and velocity measurements, which are key sensor readings for identifying and interpreting maneuvers in traffic [23, 24]. To extend BNs with support for continuous variables, research has pursued various directions, which we review in the next sub-sections.

3.1 Algorithms based on the Gaussian assumption

One may choose to carefully construct the model such that exact inference algorithms can still be applied. Computationally, this requires that the joint distribution over the variables of the domain is from a distribution-class that is closed under certain operations required for inference [25].

This is the approach followed in the algorithm introduced in [26] and revised in [27]. The algorithm is able to carry out exact inference in BNs, where the joint distribution is a conditional linear Gaussian (CLG) distribution [28]. In the CLG model, the conditional distribution of each discrete variable $X_D \in \mathbf{X}$ given its parents is a multinomial, whilst the conditional distribution of each continuous variable $X_D \in \mathbf{X}$ with discrete parents $\mathbf{X}_D \subseteq \mathbf{X}$ and continuous parents $\mathbf{X}_C \subseteq \mathbf{X}$, is given by

$$f(z|\mathbf{X}_D = \mathbf{x}_D, \mathbf{X}_C = \mathbf{x}_C) = \mathcal{N}(z; \alpha(\mathbf{x}_D) + \boldsymbol{\beta}(\mathbf{x}_D)^{\mathsf{T}}\mathbf{x}_C, \sigma^2(\mathbf{x}_D)), \tag{4}$$

for all $\mathbf{x}_D \in \Omega_{\mathbf{X}_D}$ and $\mathbf{x}_C \in \Omega_{\mathbf{X}_C}$, where α and $\boldsymbol{\beta}$ are the coefficients of a linear regression model of Z given its continuous parents; this model can differ for each configuration of the discrete variables $\mathbf{X}_D$.

After fixing any configuration of the discrete variables, the joint distribution of any subset $\mathbf{X}_C \subseteq \mathbf{X}$ of continuous variables is a multivariate Gaussian. Hence, the parameters of the multivariate Gaussian can be obtained from the ones in the CLG representation. For a set of n continuous variables $Z_1, \ldots, Z_n$ with a conditionally specified joint density

$$f(z_1, \ldots, z_n) = \prod_{i=1}^{n} f(z_i | z_{i+1}, \ldots, z_n)\,, \tag{5}$$

where the k-th factor, $1 \leq k \leq n$, is such that

$$f(z_k | z_{k+1}, \ldots, z_n) = \mathcal{N}(z_k; \mu_{z_k|z_{k+1},\ldots,z_n}, \sigma^2_{z_k|z_{k+1},\ldots,z_n})\,,$$

it holds that the joint is

$$f(z_1, \ldots, z_n) = \mathcal{N}(z_1, \ldots, z_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})\,,$$

where $\boldsymbol{\mu}$ is the n-dimensional vector of means and $\boldsymbol{\Sigma}$ is the covariance matrix of the multivariate distribution over random variables $Z_1, \ldots, Z_n$ and both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are derived from the parameters in Eq. (4) [29].

A numerically more stable algorithm that operates under the CLG framework is given in [30]. It operates with univariate regressions on the continuous variables, avoiding matrix manipulations and the possible numerical errors entailed by these manipulations.

A more efficient algorithm for exact inference in CLG networks is given in [31]. It is based on lazy propagation [5], that operates with factorised potentials in the cliques as well as in the separators of the join tree. Semantic knowledge is used to produce a refined version of the algorithm, that takes advantage of the nature of the potentials (conditional densities or not) involved in the calculations [32].

The CLG model does, however, impose certain limitations on the domain being modelled: discrete variables cannot directly depend on a continuous variable and each continuous variable must follow a conditional linear Gaussian distribution. Furthermore, inference in CLG models is in general harder than in models that are fully discrete or fully Gaussian. Even in simple models like polytree models, where exact inference is linear in the size of the model in the pure Gaussian case, approximate inference is NP-hard for CLGs [33].

The structural restrictions mentioned above are overcome in [34] by adopting the so-called *augmented CLG networks*, in which conditional densities of discrete nodes with continuous parents are modelled using the soft-max function. The drawback is that the algorithm is not able to provide exact inference results, as the product of Gaussian and soft-max functions is only approximated by a Gaussian. The scheme is similar to the method previously proposed in [35], where augmented CLG networks are also used, and

the product of a logistic and Gaussian function is approximated by a Gaussian using variational methods [36].

Augmented CLG networks are also used in [37]. The structure of the network is then modified using arc reversals in order to make it compatible with the exact algorithm in [27], until all the discrete nodes in the network are arranged so that they only have discrete parents. However, the result of the inference is approximate, as the conditional densities resulting from the arc reversal process are approximated by mixtures of Gaussians.

An approximation scheme of a different nature is followed in [38], where *general belief propagation* is utilised. The algorithm is based on loopy belief propagation with weak marginalisation, which means than only the first two moments are kept after marginalisation, instead of the full density. The posterior density for the variables of interest is computed by minimising the Kikuchi energy [39]. The algorithm operates over augmented CLG networks.

Approximate inference in CLG networks has also been approached using Monte Carlo methods. An algorithm based on adaptive *importance sampling* is proposed in [40]. Importance sampling [41] is a versatile simulation technique that in the case of inference in BNs amounts to transforming the numerator in Eq. (2) by multiplying and dividing by a distribution p^* that, unlike $p(\mathbf{x}, \mathbf{x}_E)$, is easy to handle and, more precisely, samples from it can be easily obtained. Then, $p(x_i, \mathbf{x}_E)$ can be written as

$$\begin{aligned} p(x_i, \mathbf{x}_E) &= \sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} \int_{\mathbf{x}_C \in \Omega_{\mathbf{X}_C}} p(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_C \\ &= \sum_{\mathbf{x}_D \in \Omega_{\mathbf{X}_D}} \int_{\mathbf{x}_C \in \Omega_{\mathbf{X}_C}} \frac{p(\mathbf{x}, \mathbf{x}_E)}{p^*(\mathbf{x}, \mathbf{x}_E)} p^*(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_C \\ &= E\left[\frac{p(\mathbf{x}, \mathbf{x}_E)}{p^*(\mathbf{x}, \mathbf{x}_E)}\right], \end{aligned} \tag{6}$$

where the expected value is computed with respect to p^*. Therefore, the problem is mapped to estimating an expected value, which can be attained by using a sample mean estimator.

The idea in [40] is that each variable is sampled from a conditional density given its parents in the network that is dynamically updated as the sampling process progresses. The sampling order is therefore from parents to children. The sampling conditional densities are modelled using CLG functions.

Rao-Blackwellisation is used in [42] as the core of an algorithm than combines Gibbs sampling with exact computations in a join tree. The idea is that in CLGs, inference on the continuous variables when the discrete ones are observed is tractable. Then, the posterior densities are estimated by sampling on the discrete variables and manipulating the conditional densities given the sample.

Scalability is addressed in [43]. The algorithm only provides exact results in CLG networks with polytree structure. Otherwise, mixture of Gaussian approximations are used.

3.2 General algorithms

The next group of algorithms are able to operate over general densities, and therefore do not rely on the Gaussian assumption. All of them are, however, of approximate nature. Examples of general approximation schemes include the Gibbs sampler [44, 45], variational inference [36, 46], and expectation propagation [47].

Variational inference is a deterministic approximate inference technique, where we seek to iteratively optimise a variational approximation to the posterior distribution of interest [46]. Let $\mathcal{Q}$ be the set of possible approximations; then the variational approximation to a posterior distribution $p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$ is defined as

$$q^*_{\mathbf{x}_E}(\mathbf{x}_I) = \arg\min_{q\in\mathcal{Q}} D(q(\mathbf{x}_I)||p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)),$$

where $D(q||p)$ is the KL divergence from q to p. An alternative is to focus on $D(p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)||q(\mathbf{x}_I))$, which corresponds to expectation propagation [47].

A common approach is to employ a variational mean-field approximation of the posterior distribution, so that the approximation factors over the individual variables involved, i.e., $q^*_{\mathbf{x}_E}(\mathbf{x}_I) = \prod_{i\in I} q^*_{\mathbf{x}_E}(x_i)$. During the optimisation of the variational mean-field one performs a coordinate ascent, where we iteratively update the individual variational distributions while holding the others fixed [48]. Updating a variational distribution essentially involves calculating the variational expectation of the original log conditional distributions of the model. This can be done efficiently and in closed form when the distributions involved are conjugate-exponential [49]. When the distributions fall outside this family, which is e.g. the case with logistic models, one may for instance resort to developing tight lower bounds to commonly occurring functions [36, 50] or apply stochastic optimisation for calculating the expectations [51, 52].

Even if the variational mean-field approximation shares many commonalities with expectation propagation, the two differ by the former having the interpretation that $q^*_{\mathbf{x}_E}(\mathbf{x}_I)$ is chosen as to maximise a lower-bound of the data likelihood $p(\mathbf{x}_E)$ while the latter does not. A general architecture for supporting variational message passing in graphical models is presented in [53], highlighting how distributions that are conjugate-exponential families [46, 49, 53] can be utilised to efficiently represent the messages by the expected natural statistics. A similar scheme can also be deployed for expectation propagation, but there relying on a transformation between the exponential family representation's expected sufficient statistics and the distribution's moments.

The scheme proposed in [54] is based on the Shenoy-Shafer architecture. Approximation is two-fold. First, messages over the join tree are approximated in order to reduce their complexity, and afterwards sampling is performed in each clique. The posterior densities are estimated from the sample.

Importance sampling is the methodology underlying the proposal in [55]. It is valid for any density representation that allows sampling. The posterior densities are given in terms of mixtures of Gaussians. A similar approach enhanced with the ability to deal with deterministic conditionals is given in [56].

The method introduced in [57] is an approximate algorithm that works for models within the exponential family. It is based on expectation propagation and is formulated for the general context of *factor graphs*, which contain Bayesian networks as special cases.

A general algorithm for inference in hybrid BNs was also proposed in [58]. The fundamental idea is based on arc reversal and approximation by mixture of Gaussians of the densities in the network, incorporating the ability of dealing with deterministic conditionals.

The so-called *nonparametric belief propagation* [59] also operates over general hybrid BNs. It is based on approximating densities by sampling and then fitting a kernel to the sample. The drawback is that kernels are not efficient representations for large samples or in complex models.

3.3 Algorithms based on transforming the original densities

Finally, one can "translate" the original model into an approximate model, for which efficient exact and approximate inference algorithms can be applied. This can be achieved by discretising the continuous variables. The method proposed in [60] performs the discretisation by minimising the Kullback-Leibler divergence dynamically on the join tree. This is, however, a computationally costly method that even requires a specific data structure, the so-called BSP-trees. A much more efficient approach is proposed in [61], but instead of dynamically discretising the densities in the join tree, the individual posterior densities are discretised.

A more sophisticated approach uses translations with a higher expressive power than discretisation. This includes mixtures of truncated exponentials (MTEs) [62], mixtures of polynomials (MoPs) [63] and the more recently proposed mixtures of truncated basis functions (MoTBFs) framework [64] that aims to combine several of the previously proposed frameworks.

The first algorithm specifically designed for MTEs was proposed in [65]. It is based on the Shenoy-Shafer scheme [25]. This algorithm was later extended to networks with deterministic conditionals in [66].

An approximate inference scheme was proposed in [67], where the Penniless propagation algorithm [18] is adapted to MTEs. The algorithm was refined and compared to approximate inference using Markov Chain Monte Carlo in [68].

The most recent proposal for inference in hybrid BNs with MTEs [69] is based on importance sampling with approximate pre-computation [70], where the sampling distributions are obtained following a variable elimination scheme [71] where the distributions

are represented using *mixed trees* [72].

A deep analysis of inference in hybrid BNs within the context of reliability analysis is described in [73]. The MTE approach is tested versus discretisation, variational approximations and Markov Chain Monte Carlo.

The instantiation of the Shenoy-Shafer architecture to MoPs was reported in [63]. The treatment of deterministic conditionals in this context was analysed in [74] by introducing a computation scheme based on Dirac functions.

The most general proposal for exact inference within this framework is given in [75]. The scheme is based on postponing the combination and addition of potentials as much as possible, adopting a lazy scheme [5]. It is valid for MoTBFs in general.

Practical aspects of inference in hybrid BNs with deterministic conditionals are studied in [76]. Inference with MTEs and MoPs is covered in this article, including approximation strategies using re-approximation of potentials.

The main drawback of MoTBFs in general, from the point of view of inference, is that conditional densities are defined by splitting the domain of the parents into hypercubes, which means that conditional MoTBFs are piecewise defined functions. Hence, the combination of conditional MoTBFs may result in an increase of the size of the result similar to the one produced when combining probability tables. This fact limits its efficiency compared to the CLG approach. Furthermore, MoTBFs do not belong to the exponential family of distributions, and therefore general methods like variational inference cannot be applied to them.

Another approach consists of approximating the conditional densities by kernels [77]. However, given the complexity of kernels, the idea has only been applied to restricted BN structures used in Bayesian classifiers [78].

The use of multivariate mixtures of Gaussians as a proxy has been explored in [79]. It is based on the fact that mixtures of multivariate Gaussians can approximate any density arbitrarily well in $\mathbb{R}^k$ using a sufficiently large number of components [80]. The resulting models are called *nonparametric Bayesian networks*. In comparison with CLGs, nonparametric BNs sidestep the Gaussian assumption and the linearity assumption as well, as the relation between a node and its parents is, in this case, a mixture of linear models rather than a single linear model. The underlying idea is very similar to the procedure for learning mixtures of Bayesian networks given in [81].

Probabilistic inference in hybrid models has also been studied from the perspective of other graphical models related to Bayesian networks. Of special interest are the *probabilistic decision graphs* (PDGs) [82], which are particularly designed to allow efficient inference. A convenient feature of PDGs is that any discrete BN can be transformed into a PDG [82]. It is also known that CLG models can be represented as PDGs [29] and therefore inference on such hybrid models can be efficiently performed over them. Recently, it has been shown that MoTBFs can also be represented as PDGs, though so far they have only been tested in practice in supervised classification problems [83].

4 MAP inference in hybrid Bayesian networks

The only specific work on MAP inference in hybrid BNs we have found in the literature is [84]. It describes how to apply the Hugin architecture [85] to the MAP problem in hybrid BNs, by distinguishing between sum and max marginalisation. They point out the problem resulting after max-marginalisation, that can be illustrated as follows.

- Consider a binary variable D and a Gaussian X so that the joint distribution is [84]

$$f(d,x) = \begin{cases} 0.2\mathcal{N}(x;1,0.1) & \text{if D=0,} \\ 0.8\mathcal{N}(x;3,3) & \text{if D=1.} \end{cases}$$

- Max-marginalising out D we get

$$f(x) = \max\{0.2\mathcal{N}(x;1,0.1), 0.8\mathcal{N}(x;3,3)\}.$$

A plot of density f is given in Figure 3, overlaid on its two Gaussian components.

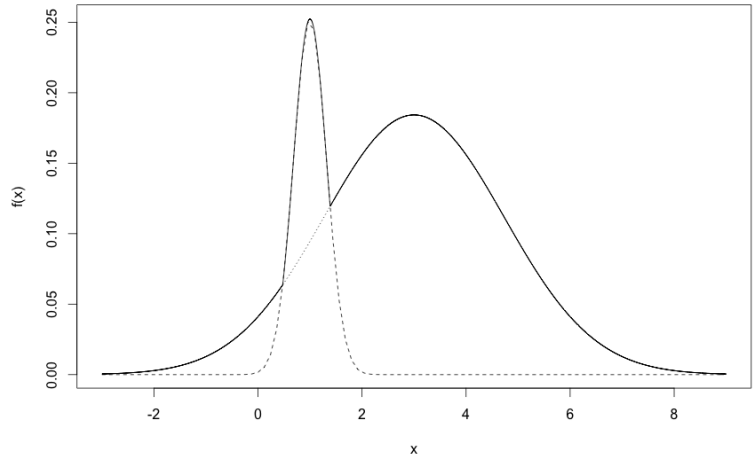

Figure 3: A plot of the maximum of two Gaussian densities.

Note that, unlike the original densities, the resulting function is actually defined by parts. This fact is an added burden with respect to inference, as handling split points significantly increases the complexity of the reasoning process.

The instantiation of the MAP problem to hybrid domains conveys an added difficulty, as it requires a maximisation of the probability density over the continuous variables, which in many cases cannot be solved analytically, as for instance when using MTEs [62].

The algorithm described in [86] is formulated in a general way and therefore valid for hybrid BNs. However, it is restricted to problems where the maximum is computed over all non-observed variables, i.e. MPE. Furthermore, convergence to the global solution is only guaranteed in the pure Gaussian case given the unimodality of the Gaussian density. However, the solution may be only a local optimum in multimodal settings like the CLG.

An important issue to take into account when considering MAP inference with continuous variables is the meaning of its output in practical problems. Consider, for instance, a scenario where there is only one variable of interest and its posterior density is a mixture of two Gaussians with identical standard deviation, as depicted in Fig. 4. In such situations, the solution to (3) may be not unique (in fact, this can also be the case in the discrete case, where the maximum probability may be reached at different configurations of values). But even if it was unique, it may be of no practical value, as a single point of a continuous variable actually represents an event of probability zero. Rather than considering a single value (or configuration in the multivariate case) one may choose to compute a *region* within the space of possible values of the variables of interest, with highest probability of occurrence according to the posterior density. Such regions are called *highest posterior density regions* [87].

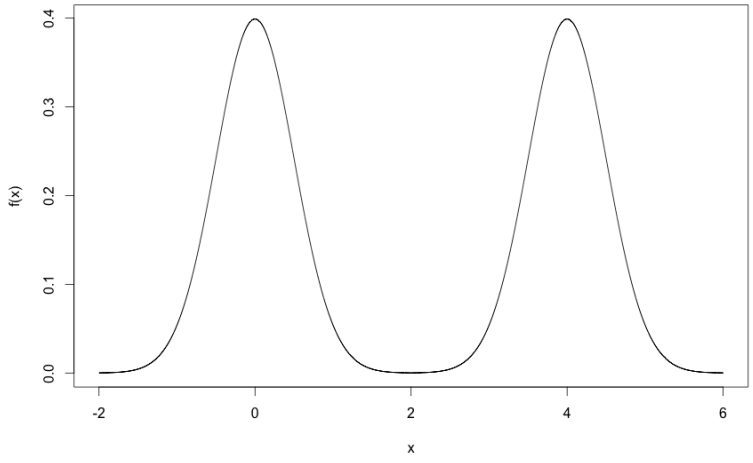

Figure 4: A plot of a mixture of two Gaussian densities with equal standard deviation.

For a posterior density $f(x)$ and a fixed value $\alpha < 1$, computing a highest posterior density region amounts to finding a and b such that

$$\alpha = \int_a^b f(x)\mathrm{d}x. \tag{7}$$

Note that in general multiple solutions to Eq. (7) are possible, and therefore some criterion must be specified, as for instance, obtaining the pair (a, b) such that the volume of the region is minimum, or locate it around the mean, median or mode.

If the MAP problem is approached as the identification of a region according to the posterior density of the variables of interest, then its relation to the plain inference problem is more tangible. More precisely, the computation of the posterior density could be carried out using the algorithms reviewed in Sect. 3.

From a practical point of view it may also be valid to discretise the continuous variables and then run a MAP inference algorithm for discrete variables. One of the first attempts to solve the MAP problem in discrete networks was based on the use of *bucket elimination* [88], similar to variable elimination. An exact algorithm based on the branch-and-bound technique was proposed in [7], where also an approximate method for computing an upper bound on the MAP configuration is given. Scalability has been addressed in [89], where a parallel algorithm is presented. However, exact solutions are not guaranteed.

5 Inference in dynamic Bayesian networks

Many domains can be seen as having strong internal structure; a feature particularly evident among those analysed in the AMIDST project [24]. This will be apparent by the domains being appropriately described using an object oriented language, either due to repetitive substructures or substructures that can naturally be ordered in a superclass–subclass hierarchy. We can find this property either in dynamic domains, where the same objects are repeated over time, or for snapshot models where many similar objects are observed simultaneously. Object oriented Bayesian networks [90, 91] (OOBNs) are defined to take advantage of such internal model structure. *Dynamic Bayesian networks* [92–94], which are used to model domains that develop over time by encoding the temporal dynamics of the system explicitly, can therefore also be represented as OOBNs.

For regular types of dynamic Bayesian networks we usually assume that observations are made at a fixed time-frequency and that the probability distribution underlying the transition model and the observation model are invariant over time, even if continuous-time Bayesian networks have also been investigated [95]. Most dynamic Bayesian networks used in practice are also assumed to follow a first order Markov process, the sensor Markov assumption, and be stationary (see, for example, the hidden Markov model (HMM) [96], which is a special type of dynamic Bayesian network). In these models, a latent (vector) variable $\mathbf{X}_t$ is used to represent the *belief state* at time t and the (vector) variable $\mathbf{E}_t$ the observations. The dynamic aspect of the model is conveyed only through the connection of the belief states over time. Smyth [97] presents a simple dynamic model of this type for detecting both known and unknown failure states of a system, and he also demonstrates the necessity of modelling temporal dynamics in order to achieve high prediction accuracy. The *Input-Output* HMMs [98] extends the HMMs by explicitly modelling "input" variables $\mathbf{U}_t$ that are controlled externally, resulting in

an effect from the "output" variables $\mathbf{E}_t$. An example of a three time slice input-output HMM is shown in Figure 5.

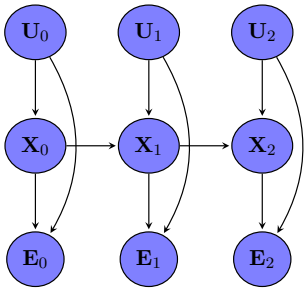

Figure 5: A three time slice input-output HMM.

Inference in dynamic Bayesian networks obviously share the computational difficulties of regular Bayesian networks, but in the dynamic case we are also often faced with additional problems. One is the entanglement problem, where after a certain time step, all variables $\mathbf{x}_t$ describing the belief state have become dependent after observing $\{\mathbf{u}_{1:t}, \mathbf{e}_{1:t}\}$, and we can therefore not represent the exact belief state $p(\mathbf{x}_t|\mathbf{e}_{1:t}, \mathbf{u}_{1:t})$ in factorised form [99, 100]. As an example, Figure 6 illustrates a dynamic Bayesian network, where all variables used to encode the belief state at time $t = 2$, i.e., X_2^1, X_2^2, and X_2^3, have become dependent after observing the evidence $\{e_0, e_1, e_2\}$. Rather than dealing with this fully correlated belief state one often employs approximate methods including approximate factorisations of the joint probability distribution describing the system state [99] as well as sampling based techniques in the form of particle filtering [101]. An important specialisation of the HMMs, which enforces internal structure on the belief state during specification, is the factorial HMMs [102], and its later extension infinite factorial HMMs [103] further generalised in [104]. Inference in the two latter model classes is performed using sampling.

Similar to the extension of the static Bayesian network model to hybrid domains, dynamic Bayesian network models have likewise been extended to continuous and hybrid domains [105]. In purely continuous domains, where the continuous variables follow linear Gaussian distributions, the dynamic Bayesian network corresponds to (a factorised version of) a Kalman filter [106]. In these types of models, the dynamics of the process are assumed to be linear, and exact inference can be performed efficiently. When modelling non-linear domains, the dynamics and observational distributions are often approximated through, e.g., the extended Kalman filter or the unscented Kalman filter [107].

In hybrid domains where the continuous variables follow a conditional linear Gaussian distribution, the model is also known as a switching dynamical linear system (SDLS), where changes in the discrete variables cause the continuous linear dynamics to switch. A particular problem with an SDLS model is that the marginal distribution over the

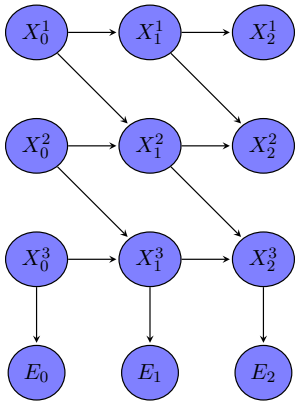

Figure 6: A three time slice dynamic Bayesian network. All variables used to describe the belief state (i.e. X_2^1, X_2^2, and X_2^3) have become correlated at time $t = 2$.

continuous variables (at a certain point in time) is a mixture of multivariate Gaussian distributions, and the number of mixture components grow exponentially and unboundedly over time. In order to ensure that the mixture of Gaussians does not grow too large several approximate solution techniques have been proposed. For example, [108] collapses a mixture of Gaussians into a single Gaussian in order to avoid the exponential blow-up in the number of mixture components. Alternative, one could also apply discretization approaches [109] or restrict the model to only cover the k most recent observations, thereby effectively limiting the number of mixture components.

Variational approximations have also been used in this setting (see, e.g., [110]), but as noted by, e.g., [111], the compactness-property of variational inference (which is a consequence of the calculation scheme's built-in preference for approximations $q(\mathbf{x})$ that are strictly positive only when the true distribution $p(\mathbf{x})$ is strictly positive) can lead to a failure to propagate uncertainty in time, thus limiting the usefulness of the calculated belief states. Alternatives include Expectation propagation [112] and structured variational Bayes [36]. The latter relaxes the mean-field approximation by assuming a factorisation of the variational approximation into factors containing more than single variables, thereby increasing the ability to represent the complexity of the posterior belief state. Unfortunately, structured variational Bayes solutions must be tailor-made to the problem at hand, as, e.g., in [102].

A Bayesian formulation of a stationary dynamic Bayesian network complicates the inference. Each model parameter, modelled as an (unobserved) latent variable, will have a child in each time-step that effectively introduces new correlations over time. These variables, denoted *global hidden variables* [52], require special attention using the variational Bayes formulation [52, 113–115].

6 Conclusions

This document describes the state-of-the art of inference in hybrid and dynamic models. While inference in static hybrid BNs has been widely studied, scalability has only received a limited attention. Given the complexity of the inference task, approximate algorithms are of special interest.

MAP inference in hybrid BNs has been only marginally studied, and still remains as an open problem even if scalability issues are not considered. An important difference with respect to the discrete case is the meaning of the outcome of a MAP algorithm, as giving a precise MAP configuration when continuous variables are involved may be useless in practical applications.

The complexity of inference in dynamic models not only resembles the difficulties of the static case, but the inclusion of the time component enlarges the magnitude of the problem and emphasises the need of approximate scalable algorithms. The problem is even more complex if a Bayesian formulation is adopted.

References

[1] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Mateo, CA. (1988)

[2] Cooper, G.F.: The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. Artificial Intelligence **42** (1990) 393–405

[3] Dagum, P., Luby, M.: Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial Intelligence **60**(1) (1993) 141–153

[4] Shafer, G.R., Shenoy, P.P.: Probability propagation. Annals of Mathematics and Artificial Intelligence **2** (1990) 327–352

[5] Madsen, A.L., Jensen, F.V.: Lazy propagation: A junction tree inference algorithm based on lazy evaluation. Artificial Intelligence **113** (1999) 203–245

[6] Darwiche, A.: Recursive conditioning. Artificial Intelligence **125**(1–2) (2001) 5–41

[7] Darwiche, A.: A differential approach to inference in Bayesian networks. Journal of the Association for Computing Machinery **50**(3) (2003) 280–305

[8] Díez, F., Mira, J.: Distributed inference in Bayesian networks. Cybernetics and Systems **25** (1994) 39–61

[9] Kozlov, A.V., Singh, J.P.: A parallel Lauritzen-Spiegelhalter algorithm for probabilistic inference. In: Proceedings of the 1994 ACM/IEEE conference on Supercomputing - Supercomputing '94, ACM Press (November 1994) 320–329

[10] Kozlov, A.V., Singh, J.P.: Parallel implementations of probabilistic inference. Computer **29**(12) (1996) 33–40

[11] Pennock, D.M.: Logarithmic time parallel Bayesian inference. In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. (July 1998) 431–438

[12] Madsen, A.L., Jensen, F.V.: Parallelization of inference in Bayesian networks. Technical Report R-99-5002, Department of Computer Science, Aalborg University (1999)

[13] Namasivayam, V., Prasanna, V.: Scalable parallel implementation of exact inference in Bayesian networks. In: 12th International Conference on Parallel and Distributed Systems - (ICPADS'06). Volume 1., IEEE (2006) 8 pp.

[14] Nikolova, O., Zola, J., Aluru, S.: A parallel algorithm for exact Bayesian network inference. In: 2009 International Conference on High Performance Computing (HiPC), IEEE (December 2009) 342–349

[15] Díez, F., Druzdzel, M.: Canonical probabilistic models for knowledge engineering. Technical Report 06-01, CISIAD (2006)

[16] Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning **29** (1997) 131–163

[17] Kjærulff, U.: Reduction of computational complexity in Bayesian networks through removal of weak dependencies. In: Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann (1994) 374–382

[18] Cano, A., Moral, S., Salmerón, A.: Penniless propagation in join trees. International Journal of Intelligent Systems **15** (2000) 1027–1059

[19] Gámez, J.: Abductive inference in Bayesian networks: a review. In Gámez, J., Moral, S., Salmerón, A., eds.: Advances in Bayesian Networks. Springer (2004) 101–117

[20] de Campos, C.P.: New complexity results for MAP in Bayesian networks. In Walsh, T., ed.: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. (2011) 2100–2106

[21] Kwisthout, J.: Most probable explanations in Bayesian networks: Complexity and tractability. International Journal of Approximate Reasoning **52** (2011) 1452–1469

[22] Choi, A., Darwiche, A.: Relax then compensate: On max-product belief propagation and more. In: Proceedings of the Twenty-Third Annual Conference on Neural Information Processing Systems (NIPS). (2009) 351–359

[23] Borchani, H., Fernández, A., Gundersen, O.E., Hovda, S., Langseth, H., Madsen, A.L., Martínez, R.S., Masegosa, A., Nielsen, T.D., Salmerón, A., Sørmo, F., Weidl, G.: Requirements for the automotive, oil and financial data domains (2014) Deliverable 1.2 of the AMIDST project, `amidst.eu`.

[24] Borchani, H., Fernández, A., Gundersen, O.E., Hovda, S., Langseth, H., Madsen, A.L., Martínez, A.M., Martínez, R.S., Masegosa, A., Nielsen, T.D., Salmerón, A., Sørmo, F., Weidl, G.: The AMIDST modelling framework – Initial draft report (2014) Deliverable 2.1 of the AMIDST project, `amidst.eu`.

[25] Shenoy, P., Shafer, G.: Axioms for probability and belief-function propagation. In: Proceedings of the Sixth Workshop on Uncertainty in Artificial Intelligence. (1990) 169–198

[26] Lauritzen, S.L.: Propagation of probabilities, means and variances in mixed graphical association models. Journal of the American Statistical Association **87**(420) (1992) 1098–1108

[27] Lauritzen, S.L., Jensen, F.: Stable local computation with conditional Gaussian distributions. Statistics and Computing **11**(2) (2001) 191–203

[28] Lauritzen, S., Wermuth, N.: Graphical models for associations between variables, some of which are qualitative and some quantitative. The Annals of Statistics **17** (1989) 31–57

[29] Nielsen, J., Gámez, J., Salmerón, A.: Modelling and inference with Conditional Gaussian probabilistic decision graphs. International Journal of Approximate Reasoning **53** (2012) 929–945

[30] Cowel, R.: Local propagation in conditional Gaussian Bayesian networks. Journal of Machine Learning Research **6** (2005) 1517–1550

[31] Madsen, A.: Belief update in CLG Bayesian networks with lazy propagation. International Journal of Approximate Reasoning **49** (2008) 503–521

[32] Zhu, M., Liu, S., Yang, Y.: Propagation in CLG Bayesian networks based on semantic modeling. Artificial Intelligence Review **38** (2012) 149–162

[33] Lerner, U., Parr, R.: Inference in hybrid networks: Theoretical limits and practical algorithms. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann (2001) 310–318

[34] Lerner, U., Segal, E., Koller, D.: Exact inference in networks with discrete children of continuous parents. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. (2001) 319–328

[35] Murphy, K.: A variational approximation for Bayesian networks with discrete and continuous latent variables. In: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. (1999) 457–466

[36] Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. Machine Learning **37** (1999) 183–233

[37] Shenoy, P.: Inference in hybrid Bayesian networks with mixtures of Gaussians. In: Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence. (2006) 428–436

[38] Heskes, T., Zoeter, O.: Generalized belief propagation for approximate inference in hybrid Bayesian networks. In: Proceedings of the 9th Workshop on A.I. and Statistics. (2003)

[39] Yedidia, J., Freeman, W., Weiss, Y.: Bethe free energy, Kikuchi approximations and belief propagation algorithms. In: Advances in Neural Information Processing Systems. Volume 13. (2001)

[40] Sun, W., Chang, K.: Probabilistic inference using linear Gaussian importance sampling for hybrid Bayesian networks. In: Signal Processing, Sensor Fusion, and Target Recognition XIV. Proc. of SPIE. Volume 5809. (2005) 322–329

[41] Hammersley, J., Handscomb, D.: Monte Carlo Methods. Chapman & Hall (1964)

[42] Paskin, M.: Sample propagation. In Thrun, S., Saul, L., Schölkopf, B., eds.: Advances in Neural Information Processing Systems 16, MIT Press (2004) 425–432

[43] Sun, W., Chang, K., Laskey, K.: Scalable inference for hybrid Bayesian networks with full density estimations. In: Proceedings of the 13th Conference on Information Fusion. (2010) 1–8

[44] Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984) 721–741

[45] Hrycej, T.: Gibbs sampling in Bayesian networks (research note). Artificial Intelligence **46** (1990) 351–363

[46] Attias, H.: A variational Bayesian framework for graphical models. Advances in neural information processing systems (2000) 209—215

[47] Minka, T.: Expectation propagation for approximate Bayesian inference. In: Proceedings of the Seventeenth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-01). (2001) 362–369

[48] Jaakkola, T.S., Qi, Y.: Parameter expanded variational Bayesian methods. In: Advances in Neural Information Processing Systems. (2006) 1097–1104

[49] Beal, M.J.: Variational algorithms for approximate Bayesian inference. PhD thesis, Gatsby Computational Neuroscience Unit, University College London (2003)

[50] Jaakkola, T.S., Jordan, M.I.: Bayesian parameter estimation via variational methods. Statistics and Computing **10**(1) (January 2000) 25–37

[51] Paisley, J.W., Blei, D.M., Jordan, M.I.: Variational Bayesian inference with stochastic search. In: ICML. (2012)

[52] Hoffman, M., Blei, D., Wang, C., Paisley, J.: Stochastic variational inference. The Journal of Machine Learning Research **14** (2013) 1303–1347

[53] Winn, J., Bishop, C.: Variational message passing. Journal of Machine Learning Research **6** (2005) 661–694

[54] Koller, D., Lerner, U., Anguelov, D.: A general algorithm for approximate inference and its application to hybrid Bayes nets. In: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. (1999) 324–333

[55] Yuan, C., Druzdzel, M.: Importance sampling for general hybrid Bayesian networks. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. (2007) 652–659

[56] Yuan, C., Druzdzel, M.: Generalized evidence pre-propagated importance sampling for hybrid Bayesian networks. In: AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence. Volume 2. (2007) 1296–1302

[57] Heskes, T., Opper, M., Wiegerinck, W., Winther, O., Zoeter, O.: Approximate inference techniques with expectation constraints. Journal of Statistical Mechanics: Theory and Experiment (2005) 1105

[58] Cinicioglu, E., Shenoy, P.: Arc reversals in hybrid Bayesian networks with deterministic variables. International Journal of Approximate Reasoning **50** (2009) 763–777

[59] Sudderth, E., Ihler, A., Isard, M., Freeman, W., Willsky, A.: Nonparametric belief propagation. Communications of the ACM **53**(10) (2010) 95–103

[60] Kozlov, A.V., Koller, D.: Nonuniform dynamic discretization in hybrid networks. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. (1997) 314–325

[61] Neil, M., Tailor, M., Marquez, D.: Inference in Bayesian networks using dynamic discretisation. Statistics and Computing **17**(3) (2007) 219–233

[62] Moral, S., Rumí, R., Salmerón, A.: Mixtures of truncated exponentials in hybrid Bayesian networks. In: EQSCARU'2001. Volume 2143 of Lecture Notes in Artificial Intelligence. Springer, Berlin, Germany (2001) 145–167

[63] Shenoy, P., West, J.: Inference in hybrid Bayesian networks using mixtures of polynomials. International Journal of Approximate Reasoning **52** (2011) 641–657

[64] Langseth, H., Nielsen, T.D., Rumí, R., Salmerón, A.: Mixtures of truncated basis functions. International Journal of Approximate Reasoning **53**(2) (2012) 212–227

[65] Cobb, B., Shenoy, P.: Inference in hybrid Bayesian networks with mixtures of truncated exponentials. International Journal of Approximate Reasoning **41** (2006) 257–286

[66] Cobb, B., Shenoy, P.: Operations for inference in continuous Bayesian networks with linear deterministic variables. International Journal of Approximate Reasoning **42** (2006) 21–36

[67] Rumí, R., Salmerón, A.: Penniless propagation with mixtures of truncated exponentials. Lecture Notes in Computer Science **3571** (2005) 39–50

[68] Rumí, R., Salmerón, A.: Approximate probability propagation with mixtures of truncated exponentials. International Journal of Approximate Reasoning **45** (2007) 191–210

[69] Fernández, A., Rumí, R., Salmerón, A.: Answering queries in hybrid Bayesian networks using importance sampling. Decision Support Systems **53** (2012) 580–590

[70] Salmerón, A., Cano, A., Moral, S.: Importance sampling in Bayesian networks using probability trees. Computational Statistics and Data Analysis **34** (2000) 387–413

[71] Zhang, N., Poole, D.: Exploiting causal independence in Bayesian network inference. Journal of Artificial Intelligence Research **5** (1996) 301–328

[72] Moral, S., Rumí, R., Salmerón, A.: Approximating conditional MTE distributions by means of mixed trees. ECSQARU'03. Lecture Notes in Artificial Intelligence **2711** (2003) 173–183

[73] Langseth, H., Nielsen, T., Rumí, R., Salmerón, A.: Inference in hybrid Bayesian networks. Reliability Engineering and System Safety **94** (2009) 1499–1509

[74] Shenoy, P.P., West, J.C.: Extended Shenoy-Shafer architecture for inference in hybrid Bayesian networks with deterministic conditionals. International Journal of Approximate Reasoning **52**(6) (2011) 805–818

[75] Langseth, H., Nielsen, T., Rumí, R., Salmerón, A.: Inference in hybrid Bayesian networks with mixtures of truncated basis functions. In: Proceedings of the Sixth European Workshop on Probabilistic Graphical Models (PGM'2012. (2012) 171–178

[76] Shenoy, P., Rumí, R., Salmerón, A.: Practical aspects of solving hybrid Bayesian networks containing deterministic conditionals. International Journal of Intelligent Systems (2014) (To appear)

[77] Wand, M., Jones, M.: Kernel smoothing. Chapman & Hall, London (1995)

[78] Pérez, A., Larrañaga, P., Inza, I.: Bayesian classifiers based on kernel density estimation: Flexible classifiers. International Journal of Approximate Reasoning **50** (2009) 341–362

[79] Ickstadt, K., Bornkamp, B., Grzegorczyk, M., Wieczorek, J., Sheriff, M.R., Grecco, H., Zamir, E.: Nonparametric Bayesian networks. In Bernardo, J., Bayarri, M., Berger, J., Heckerman, A.D.D., Smith, A., West, M., eds.: Bayesian Statistics. Volume 9. Oxford University Press (2010)

[80] Wu, Y., Ghosal, S.: Kullback Leibler property of kerner mixture priors in Bayesian density estimation. Electronic Journal of Statistics **2** (2008) 298–331

[81] Thiesson, B., Meek, C., Chickering, D., Heckerman, D.: Learning mixtures of Bayesian networks. In Cooper, G., Moral, S., eds.: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. (1998)

[82] Jaeger, M.: Probabilistic decision graphs – combining verification and AI techniques for probabilistic inference. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **12** (2004) 19–42

[83] Fernández, A., Rumí, R., del Sagrado, J., Salmerón, A.: Supervised classification using hybrid probabilistic decision graphs. In van der Gaag, L., Feelders, A., eds.: PGM 2014. Volume 8754 of Lecture Notes in Artificial Intelligence., Springer (2014) 206–221

[84] Sun, W., Chang, K.: Study of the most probable explanation in hybrid Bayesian networks. In: Signal Processing, Sensor Fusion, and Target Recognition XX. Proc. of SPIE. Volume 8050. (2011)

[85] Jensen, F., Lauritzen, S., Olesen, K.: Bayesian updating in causal probabilistic networks by local computation. Computational Statistics Quarterly **4** (1990) 269–282

[86] Weiss, Y., Freeman, W.: On the optimality of solutions of the Max-Product belief-propagation algorithm in arbitrary graphs. IEEE Transactions on Information Theory **47**(2) (2010) 736–744

[87] Turkkan, N., Pham-Gia, T.: Highest posterior density credible region and minimum area confidence region: the bivariate case. Applied Statistics **46**(1) (1997) 131–182

[88] Dechter, R.: Bucket elimination: a unifiying framework for reasoning. Artificial Intelligence **113** (1999) 41–85

[89] Fu, Q., Wang, H., Banerjee, A.: Bethe-ADMM for tree decomposition based parallel MAP inference. In: Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI2013). (2013)

[90] Koller, D., Pfeffer, A.: Object-oriented Bayesian networks. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann (1997) 302–313

[91] Neal, M., Fenton, N., Nielson, L.: Building large-scale Bayesian networks. The Knowledge Engineering Review **15**(3) (2000) 257–284

[92] Dean, T., Kanazawa, K.: A model for reasoning about persistence and causation. Computational Intelligence **5**(3) (1989) 142–150

[93] Kjærulff, U.: A computational scheme for reasoning in dynamic probabilistic networks. In: Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence, San Francisco, California, Morgan Kaufmann Publishers (1992) 121–129

[94] Murphy, K.P.: Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, UC Berkeley, Computer Science Division (2002)

[95] Nodelman, U., Shelton, C., Koller, D.: Continuous time Bayesian networks. In: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI). (2002) 378–387

[96] Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE **77**(2) (1989) 257–286

[97] Smyth, P.: Markov monitoring with unknown states. IEEE Journal of Selected Areas in Communications, Special Issue on Intelligent Signal Processing for Communications **12**(9) (1994) 1600–1612

[98] Bengio, Y.: Markovian models for sequential data. Neural Computing Surveys (1999) 129–162

[99] Boyen, X., Koller, D.: Tractable inference for complex stochastic processes. In: Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-98). (1998) 33–42

[100] Boyen, X., Koller, D.: Exploiting the architecture of dynamic systems. In: Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99). (1999) 313–320

[101] Doucet, A., De Freitas, N., Murphy, K., Russell, S.: Rao-Blackwellised particle filtering for dynamic Bayesian networks. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence. (2000) 176–183

[102] Ghahramani, Z., Jordan, M.I.: Factorial hidden Markov models. Machine Learning **29**(2-3) (1997) 245–273

[103] Van Gael, J., Teh, Y.W., Ghahramani, Z.: The infinite factorial hidden Markov model. In: Advances in Neural Information Processing Systems. Volume 21. (2009)

[104] Doshi, F., Wingate, D., Tenenbaum, J.B., Roy, N.: Infinite dynamic Bayesian networks. In: International Conference on Machine Learning (ICML-11). (2011) 913–920

[105] Lerner, U., Moses, B., Scott, M., McIlraith, S., Koller, D.: Monitoring a Complex Physical System using a Hybrid Dynamic Bayes Net. In: Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI-02). (2002)

[106] Kalman, R.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering **82** (1960) 35–45

[107] Julier, S.J., Uhlmann, J.K.: A new extension of the Kalman filter to nonlinear systems. In: SPIE AeroSense Symposium. (1997) 182–193

[108] Barber, D.: Expectation correction for smoothed inference in switching linear dynamical systems. Journal of Machine Learning Research **7** (2006) 2515–2540

[109] Vinh, N., Chetty, M., Coppel, R., Wangikar, P.: Data discretization for dynamic bayesian network based modeling of genetic networks. In Huang, T., Zeng, Z., Li, C., Leung, C., eds.: Neural Information Processing. Volume 7664 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 298–306

[110] Ghahramani, Z., Hinton, G.E.: Variational Learning for Switching State-Space Models. Neural Computation **12**(4) (2000) 831–864

[111] Turner, R.E., Sahani, M.: Two problems with variational expectation maximisation for time-series models. In Barber, D., Cemgil, T., Chiappa, S., eds.: Bayesian Time series models. Cambridge University Press (2011) 109–130

[112] Zoeter, O., Heskes, T.: Expectation propagation and generalised EP methods for inference in switching Kalman filter models. In Barber, D., Cemgil, A.T., Chiappa, S., eds.: Probabilistic Methods for Time-Series Analysis. Cambridge University Press (2011) 181–207

[113] Ghahramani, Z., Beal, M.J.: Propagation algorithms for variational Bayesian learning. In Leen, T.K., Dietterich, T.G., Tresp, V., eds.: Advances in Neural Information Processing Systems 12, MIT Press (2000) 507–513

[114] Broderick, T., Boyd, N., Wibisono, A., Wilson, A.C., Jordan, M.: Streaming variational Bayes. In: Advances in Neural Information Processing Systems 26. Neural Information Processing Systems (2013) 1727–1735

[115] Luttinen, J.: Fast variational Bayesian linear state-space model. In: Machine Learning and Knowledge Discovery in Databases. Volume 8188 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 305–320