

A Framework for Solving Hybrid Influence Diagrams Containing Deterministic Conditional Distributions

Yijing Li, Prakash P. Shenoy

University of Kansas School of Business, Lawrence, Kansas 66045
{yjl@ku.edu, pshenoy@ku.edu}

We describe a framework and an algorithm for approximately solving a class of hybrid influence diagrams (IDs) containing discrete and continuous chance variables, discrete and continuous decision variables, and deterministic conditional distributions for chance variables. A conditional distribution for a chance variable is said to be deterministic if its variances, for each state of its parents, are all zeroes. The solution algorithm is an extension of Shenoy's fusion algorithm for discrete influence diagrams. To mitigate the integration and optimization problems associated with solving hybrid IDs, we propose using mixture of polynomials approximations of conditional probability density and utility functions and piecewise linear approximations of nonlinear deterministic conditional distributions for continuous chance variables. The class of hybrid IDs that can be solved by our framework are those that do not involve divisions. The framework and algorithm are illustrated by solving two small examples of hybrid IDs.

Key words: solving hybrid influence diagrams; deterministic conditional distributions; mixture of polynomials

History: Received on August 2, 2011. Accepted on January 14, 2012, after 1 revision.

1. Introduction

An influence diagram (ID) is a formal compact representation of a Bayesian decision making under uncertainty problem. It consists of four parts: a sequence of decisions, a set of chance variables with a joint distribution represented by a Bayesian network (BN), the decision maker's preferences for the uncertain outcomes represented by a joint utility function, and information constraints that indicate what uncertainties are known and unknown when a decision has to be made. IDs were initially defined by Howard and Matheson (1984, 2005). Howard and Matheson's (1984, 2005) definition of an ID allowed a single (unfactored) utility node. Tatman and Shachter (1990) subsequently generalized IDs to include multiple utility nodes that combine additively or multiplicatively or some combination of the two. In this paper, we assume that the utility factors combine additively.

Hybrid IDs are IDs containing a mix of discrete and continuous chance variables, and discrete and continuous decision variables. A conditional distribution (or conditional, in short) for a chance variable in an ID is said to be *deterministic* if the variances, for each state

of the variable's parents, are all zeroes. Deterministic conditionals for discrete chance variables pose no computational problems. Deterministic conditionals for continuous chance variables pose a computational challenge as the joint density function for all continuous variables does not exist, and this nonexistence can pose problems when solving such IDs. Therefore, from here onward, when we speak of variables with deterministic conditionals, we are referring to continuous variables.

In practice, one encounters decision problems in which some chance and decision variables (such as demand, cost, stock price, profit, etc.) are continuous in nature. If we maintain the continuous nature of these variables (i.e., we do not discretize such variables), the result is a hybrid ID. However, solving a hybrid ID involves two main computational challenges. First, marginalizing a continuous chance variable involves integration of a product of density and utility functions. In some cases, such as the Gaussian density function, there may not exist a closed-form representation of the integral. We will refer to this problem as the *integration* problem.

Second, marginalizing a decision variable involves maximizing a utility function. If a decision variable is continuous and has relevant continuous information predecessors, then we may be faced with the problem of finding a closed-form solution to the maximization problem. Not only do we have to find an optimal value of the decision variable as a function of the states of its relevant information predecessors, we also have to find a closed-form expression of the maximum utility as a function of the states of its relevant information predecessors. We will refer to this problem as the *optimization* problem.

In this paper, we describe a framework and an algorithm for solving a class of hybrid IDs approximately. The framework is an extension of the Shenoy and Shafer (1990) architecture for making inferences in hybrid BNs described by Shenoy and West (2011a) and includes decision variables and utility functions. The algorithm consists of using mixtures of polynomials (MOPs) for approximating probability density functions (PDFs) of continuous variables, approximating nonlinear deterministic conditionals by piecewise linear ones, using Dirac delta functions to represent deterministic conditionals for continuous chance variables, and approximating utility functions by MOPs. The class of hybrid IDs that can be solved by our framework are those IDs that can be solved using local computation without the use of the division operation. We illustrate our method by solving two small examples.

An outline of the remainder of this paper is as follows. In §2, we review the literature on solving hybrid IDs, we list the contributions of our paper, and we sketch the limitations of our method. In §3, we describe a framework and an algorithm to solve hybrid IDs with deterministic variables. In §4, we define MOP functions, a process for approximating conditional PDFs and utility functions by MOP functions, and a process for finding piecewise linear approximations of nonlinear deterministic conditionals. In §5, we solve two decision problems to illustrate our framework and algorithm. Finally, in §6, we conclude with a summary and a discussion on the limitations of using MOP functions for solving hybrid IDs and some related topics for future work.

2. Previous Work on Solving Hybrid IDs

In this section, we review previous work on solving hybrid IDs and discuss the main contributions and limitations of our method.

2.1. Discretization

A traditional method for solving a hybrid ID is to approximate the hybrid ID with a discrete ID by discretizing the continuous chance and decision variables (see, e.g., Miller and Rice 1983, Keefer and Bodily 1983, Smith 1993). If we discretize a continuous variable using too few bins, we may have an unacceptable approximation of the problem. On the other hand, if we use many bins, we increase the computational effort of solving the resulting discrete ID. In the BN literature, Kozlov and Koller (1997) described a dynamic nonuniform discretization technique for chance variables depending on the region where the posterior density lies. This technique needs to be adapted for solving hybrid IDs.

2.2. Monte Carlo Methods

Another method for solving hybrid IDs is to use Monte Carlo (MC) methods. One of the earliest to suggest MC methods for solving decision trees was Hertz (1964), which sampled the entire joint distribution of all chance variables. Charnes and Shenoy (2004) proposed an MC method that samples from a small set of chance variables at a time for each decision variable. Ortiz and Kaelbling (2000) proposed several MC methods and provided bounds on the number of samples required given some error bounds. Bielza et al. (1999) explored the use of Markov chain MC methods to solve a single-stage decision problem with continuous decision and chance nodes to solve the maximization problem. Cano et al. (2006) described a forward-backward Monte Carlo method for approximate solutions of IDs. While Monte Carlo methods can handle continuous chance variables, there is one limitation. If we have a decision variable with continuous chance variables as relevant predecessors, then finding an optimal decision function for the decision variable requires discretization of the continuous chance variables that are in the relevant domain.

2.3. Gaussian IDs

Among exact methods, Shachter and Kenley (1989) provided a theory to solve IDs where all chance and decision variables are continuous. The continuous chance variables are required to have the conditional linear Gaussian (CLG) distributions, and the utility function is required to be quadratic. Such IDs are called Gaussian IDs. These requirements ensure that the joint distribution of all chance variables is multivariate Gaussian, whose marginals can be easily found without the need for integration. Also, the quadratic nature of the utility function ensures that there is a unique maximum that can be computed in closed form without the need for searching for an optimal solution.

2.4. Mixture of Gaussian IDs

Poland (1994) and Poland and Shachter (1993) extend Gaussian IDs to include discrete chance variables that do not have continuous parents. If a continuous chance variable does not have a CLG distribution, then it can be approximated by a mixture of Gaussians represented by a discrete variable with mixture weights and a continuous variable with the discrete variable as its parent and with CLG distributions. Like Gaussian IDs, mixtures of Gaussian IDs are required to have quadratic utility functions.

2.5. Mixture of Truncated Exponentials

To find posterior marginals in hybrid BNs, Moral et al. (2001) proposed approximating PDFs by mixtures of truncated exponentials (MTEs) as a solution for the integration problem. The family of MTE functions is easy to integrate, is closed under combination and marginalization, and can be propagated using the Shenoy and Shafer (1990) architecture. Cobb et al. (2006) described MTE approximations for several commonly used univariate PDFs such as normal, log-normal, Gamma, etc. Cobb and Shenoy (2005a) extended the MTE BN framework to include one-dimensional deterministic conditionals described by linear functions. For one-dimensional nonlinear functions, Cobb and Shenoy (2005b) proposed approximating them by piecewise linear functions.

For solving IDs, Cobb and Shenoy (2008) described MTE IDs, where the PDFs of continuous chance variables and the utility functions are described using

MTE functions, and decision nodes are all discrete. Thus, any PDF can be used as long as they can be approximated by MTEs, and discrete variables can have continuous parents. Cobb (2007) described continuous decision MTE IDs, where in addition to using MTE potentials to represent PDFs and utility functions, continuous decisions are allowed.

The MTE methods surveyed here for BNs and IDs cannot cope with multidimensional linear deterministic conditionals. For example, if X and Y are independent exponential random variables with Poisson rate parameter $\lambda = 1$ (whose PDFs are MTE functions), then $Z = X + Y$ has a Gamma distribution (with parameters $r = 2$ and $\lambda = 1$), whose PDF ($f_Z(z) = z e^{-z}$ if $z > 0$) is not an MTE function (because of the presence of the $z e^{-z}$ term in the PDF).

2.6. Mixture of Polynomials.

Similar to MTEs, Shenoy and West (2011b) and Shenoy (2012) propose approximating PDFs by piecewise polynomial functions called mixtures of polynomials. Like MTEs, MOPs are closed under multiplication, addition, and integration. Thus, they can be used to find marginals in hybrid BNs using the Shenoy and Shafer (1990) architecture. MOP functions have some advantages over MTE functions. MOP approximations can be found (more easily than MTE) using Lagrange interpolating polynomials with Chebyshev points (Shenoy 2012), even for multidimensional ones. Also, they are closed for a larger class of deterministic functions than MTE functions, which are closed only for one-dimensional linear functions (e.g., $W = aX + b$). MOP functions are closed under transformations required for multidimensional linear (e.g., $W = X + Y$) and for multidimensional quotient (e.g., $W = X/Y$, $W = (X/Y)/Z$, etc.) deterministic functions.

2.7. Contributions

The major contributions of this paper are as follows. First, we further extend the extended Shenoy–Shafer architecture, described in Shenoy and West (2011a) for inference in hybrid BNs to enable the solution of hybrid IDs with deterministic conditionals. We extend the architecture to include discrete and continuous decision variables and utility functions. The algorithm for solving hybrid IDs is essentially the same as the fusion algorithm proposed by Shenoy (1992) for discrete IDs.

Second, to address the integration and optimization problems, we propose using MOP approximations of PDFs and utility functions. The family of MOP functions is closed under multiplication, addition, integration, and transformations needed for multidimensional linear deterministic functions. It is not closed under divisions or transformations needed for nonlinear deterministic functions (such as $Z = X \cdot Y$, $Y = X^2$, etc.) For hybrid IDs that contain nonlinear deterministic conditionals, we propose approximating these by piecewise linear functions as suggested by Cobb and Shenoy (2005b).

Regarding the optimization problem, because MOP functions are easily differentiable, finding the maximum of a utility function that is in MOP form is also easier than for non-MOP utility functions.

Previous methods for solving IDs containing continuous chance variables assume either CLG conditionals, in which case one can allow deterministic conditionals described by linear functions (Shachter and Kenley 1989, Poland and Shachter 1993), or non-CLG conditionals that are approximated by MTEs (Cobb and Shenoy 2008), which are closed only for one-dimensional linear deterministic conditionals (Shenoy et al. 2011). The framework described here extends the class of IDs that can be solved—the chance variables can have any distributions as long as they can be approximated by MOPs, the utility functions can be of any form as long as they can be approximated by MOPs, there are no topological restrictions such as discrete variables with no continuous parents, and we can have any deterministic conditionals as long as they can be approximated by piecewise linear functions.

2.8. Limitations

Some limitations of our method are as follows. First, the family of MOP functions is not closed under the division operation. Solving an ID with an additive factorization of the utility function using local computation may require divisions. Such problems will not be amenable to our method. The *Pigs* problem, discussed by Lauritzen and Nilsson (2001), is an example of a problem of this type (requires divisions for solution using local computation).

Second, for IDs containing deterministic conditionals, MOPs are closed only for multidimensional linear

and quotient functions. For multidimensional deterministic conditionals that are described by functions that are neither linear nor quotient, the family of MOPs is not closed under transformations required for such functions. However, if such deterministic functions can be approximated by piecewise linear ones, then one can still solve such problems using our method.

Third, because our method uses MOPs to approximate PDFs and utility functions, it inherits all the limitations of MOP-based methods. For example, finding a MOP approximation of a high-dimensional conditional PDF can be difficult. Thus, if we have a continuous chance variable with many continuous chance parents, this will pose a problem for finding a MOP approximation. Shenoy (2012) describes a MOP approximation of a three-dimensional CLG PDF. In this paper, we describe a procedure for finding a MOP approximation of a PDF using Lagrange interpolating polynomials with Chebyshev points. Using this procedure, we can find MOP approximations of the two-dimensional conditional log-normal PDFs needed to solve the *American put option* problem described by §5.2. In any case, we are not at a stage where one can fully automate the procedure of finding MOP approximations of conditional PDFs and utility functions.

3. A Framework for Solving Hybrid IDs

In this section, we describe a framework and an algorithm for solving hybrid IDs with deterministic conditionals. The framework described here is a further extension of the extended Shenoy–Shafer architecture described by Shenoy and West (2011a) for inference in hybrid BNs with deterministic conditionals. Here, we include decision variables and utility potentials, and we keep track of the nature of potentials (discrete, continuous, or utility) by keeping track of their units during the combination and the marginalization operations. The algorithm described is adapted from Shenoy (1992) for the case of discrete IDs.

3.1. Variables and States

We are concerned with a finite set $\mathcal{V} = \mathcal{D} \cup \mathcal{C}$ of variables. Variables in \mathcal{D} are called *decision* variables, and variables in \mathcal{C} are called *chance* variables. Each

variable $X \in \mathcal{V}$ is associated with a set Ω_X of possible states. If Ω_X is finite or countable, we say X is *discrete*, otherwise X is *continuous*. We will assume that the state space of continuous variables is the set of real numbers (or some measurable subset of it), and that the state space of discrete variables is a set of symbols (not necessarily real numbers). If $r \subseteq \mathcal{V}$, $r \neq \emptyset$, then $\Omega_r = \times \{\Omega_X \mid X \in r\}$. If $r = \emptyset$, we will adopt the convention that $\Omega_\emptyset = \{\diamond\}$.

We will distinguish between discrete and continuous chance variables. Let \mathcal{C}_d and \mathcal{C}_c denote the set of all discrete and continuous chance variables, respectively. Then, $\mathcal{C} = \mathcal{C}_d \cup \mathcal{C}_c$. We do not distinguish between discrete and continuous decision variables.

In an ID, each chance variable has a conditional distribution function for each state of its parents. A conditional distribution function associated with a chance variable is said to be *deterministic* if its variances (for each state of its parents) are all zeros. For example, suppose P (profit), R (revenue), and C (cost) are three continuous chance variables, and suppose R and C are parents of P . Furthermore, suppose the conditional of P is as follows: $P \mid (r, c) = r - c$ with probability 1. In this example, the conditional for P is deterministic, and we will denote it by the equation $P = R - C$.

In an ID, we will depict decision variables by rectangular nodes, discrete chance variables by single-bordered elliptical nodes, continuous chance variables with nondeterministic conditionals by double-bordered elliptical nodes, continuous chance

variables with deterministic conditionals by triple-bordered elliptical chance nodes, and additive factors of the joint utility function by diamond-shaped nodes. We do not distinguish between discrete and continuous decision variables.

An example of a hybrid ID is shown in Figure 1. This ID is a representation of the entrepreneur's problem, which will be described later in this section. In this ID, Z_1 and Z_2 are continuous chance nodes with nondeterministic conditionals; Q_n , Q_a , C_n , and C_a are continuous chance nodes with deterministic conditionals; P is a continuous decision node; and π is a utility node.

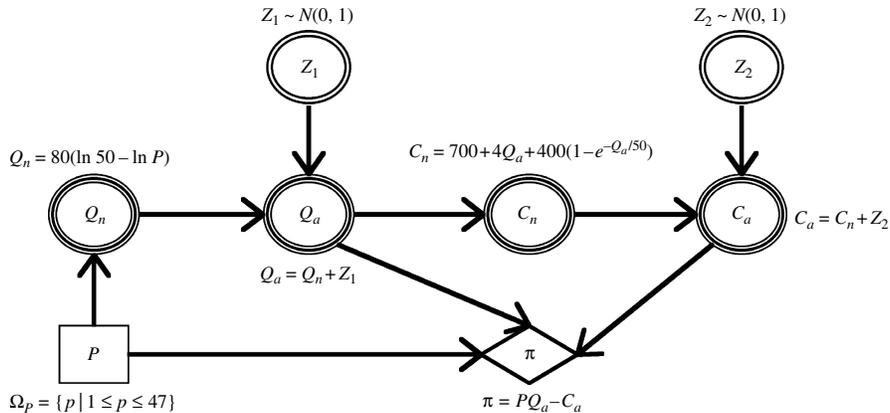
3.2. Projection of States

If $\mathbf{x} \in \Omega_r$, $\mathbf{y} \in \Omega_s$, and $r \cap s = \emptyset$, then $(\mathbf{x}, \mathbf{y}) \in \Omega_{r \cup s}$. Thus, $(\mathbf{x}, \diamond) = \mathbf{x}$. Suppose $\mathbf{x} \in \Omega_r$, and $s \subseteq r$. Then, the *projection* of \mathbf{x} to s , denoted by $\mathbf{x}^{\downarrow s}$, is the state of s obtained from \mathbf{x} by dropping states of $r \setminus s$. Thus, e.g., $(w, x, y, z)^{\downarrow \{W, X\}} = (w, x)$, where $w \in \Omega_W$, and $x \in \Omega_X$. If $s = r$, then $\mathbf{x}^{\downarrow s} = \mathbf{x}$. If $s = \emptyset$, then $\mathbf{x}^{\downarrow s} = \diamond$.

3.3. Discrete Potentials

In an ID, the conditional probability functions associated with chance variables are represented by functions called potentials. If A is discrete, it is associated with a conditional probability mass function. The conditional probability mass functions are represented by functions called discrete potentials. Formally, suppose $r \subseteq \mathcal{V}$. A *discrete potential* for r is a function $\alpha: \Omega_r \rightarrow [0, 1]$ such that the values (in the interval $[0, 1]$) are

Figure 1 An Influence Diagram Representation of the Entrepreneur's Problem



in units of probability, which are dimensionless numbers without any physical units (such as feet, pounds, seconds, etc.).

Although the domain of the potential α is Ω_r , we will refer to r as the *domain* of α . Thus, the domain of a potential representing the conditional probability function associated with some chance variable X in an ID is always the set $\{X\} \cup pa(X)$, where $pa(X)$ denotes the set of parents of X in the ID graph.

The values of discrete potentials are always in units of probability. For example, suppose B is a discrete chance variable with states b (buyer) and nb (no buyer), suppose P (price in dollars per bushel) is a continuous variable, and suppose β is a discrete potential for $\{B, P\}$, representing the conditional for B given P , such that $\beta(b, p) = 1/(1 + e^{-6.5+p})$, and $\beta(nb, p) = e^{-6.5+p}/(1 + e^{-6.5+p})$. The values of β are in units of probability.

3.4. Continuous Potentials

Continuous chance variables with nondeterministic conditionals are associated with conditional PDFs. Conditional PDFs are represented by functions called continuous potentials. Formally, suppose $r \subseteq \mathcal{V}$. A *continuous potential* ζ for r is a function $\zeta: \Omega_r \rightarrow \mathbb{R}^+$, where \mathbb{R}^+ is the set of nonnegative real numbers with units of (probability) density.

The values of continuous potentials are always in units of density. For example, suppose Y is a continuous variable whose states are in units of, say, unit Y , with continuous chance variable X as a parent. Suppose that the conditional associated with $Y | X$ is $N(x, 1)$. Then, the values of the continuous potential ψ for $\{X, Y\}$ such that $\psi(x, y) = (1/\sqrt{2\pi})e^{-(y-x)^2/2}$ are in units of probability per unit of Y , which is denoted simply by $(\text{unit}Y)^{-1}$.

Continuous variables with deterministic conditionals have conditionals described by equations. We will represent such conditionals by continuous potentials that use Dirac delta functions δ defined by Dirac (1927).

3.5. Dirac Delta Functions

A function $\delta: \mathbb{R} \rightarrow \mathbb{R}^+$ is called a *Dirac delta function* if $\delta(x) = 0$ if $x \neq 0$ and $\int_{-\infty}^{\infty} \delta(x) dx = 1$. The values of δ are in units of density.

The Dirac delta function δ is not a proper function because the value of the function at 0 doesn't 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 σ^2 in the limit as $\sigma \rightarrow 0$). However, it can be used as if it were a proper function for practically all our purposes without getting incorrect results.

Although the value $\delta(0)$ (in units of density) is undefined, i.e., ∞ , we argue that we can *interpret* the value $\delta(0)$ as probability 1 at the location $x = 0$. Consider the Gaussian PDF with mean 0 and variance σ^2 . Its moment-generating function (MGF) is $M(t) = e^{\sigma^2 t^2/2}$. In the limit as $\sigma \rightarrow 0$, $M(t) = 1$. Now, $M(t) = 1$ is the MGF of the degenerate probability distribution $X = 0$ with probability 1. Thus, we can interpret the value $\delta(0)$ as probability 1 at the location $x = 0$.

Some basic properties of the Dirac delta function are given in the appendix. An example of a deterministic conditional is as follows. Suppose R (revenue in $m\$$) is a continuous chance variable with continuous chance parents P (price in dollars per bushel) and C (crop size in $mbushels$), and discrete chance parent B with states b (buyer) and nb (no buyer). Suppose R is associated with a deterministic conditional as follows: $R = P \cdot C$ if $B = b$, and $R = 0$ if $B = nb$. Then this conditional is represented by a continuous potential ρ for $\{P, C, B, R\}$ such that $\rho(p, c, b, r) = \delta(r - p \cdot c)$, and $\rho(p, c, nb, r) = \delta(r)$. The values of ρ are in units of $(m\$)^{-1}$.

In general, if Y is a continuous variable with continuous parents $\{X_1, \dots, X_n\}$ and discrete parents $\{A_1, \dots, A_m\}$ and has a deterministic conditional $Y = g_i(X_1, \dots, X_n)$ if $(A_1, \dots, A_m) = \mathbf{a}_i$, for $i = 1, \dots, |\Omega_{\{A_1, \dots, A_m\}}|$, then such a deterministic conditional is represented by the continuous potential $\psi(\mathbf{x}, \mathbf{a}_i, y) = \delta(y - g_i(\mathbf{x}))$ for all $\mathbf{x} \in \Omega_{\{X_1, \dots, X_n\}}$, $\mathbf{a}_i \in \Omega_{\{A_1, \dots, A_m\}}$, $i = 1, \dots, |\Omega_{\{A_1, \dots, A_m\}}|$, and $y \in \Omega_Y$. The units of values of ψ are $(\text{unit}Y)^{-1}$.

3.6. Constraint Potentials

In some problems, there may be constraints on the possible states of decision variables based on states of other preceding variables. Such constraints are represented by potentials called constraint potentials. Suppose s is a set of variables such that it includes a decision variable, say, X . A *constraint potential* χ for s associated with $X \in s$ is a function $\chi: \Omega_s \rightarrow \{0, 1\}$ such that $\chi(x, \mathbf{y}) = 1$ if $x \in \Omega_X$ is a possible

alternative given $\mathbf{y} \in \Omega_{s \setminus \{X\}}$, and $\chi(x, \mathbf{y}) = 0$ if not. We assume that the constraint potential is formally specified for all states of s . In practice, it is sufficient to just specify the states of s that are possible (with the rest assumed to be not possible). The values (0 or 1) of constraint potentials are in dimensionless units. Constraint potentials are used during the process of marginalizing a decision variable.

The entrepreneur's problem discussed in §5 has a constraint on the price variable $1 \leq p \leq 47$. This constraint is handled implicitly because we expect from the nature of the problem to find an optimal price that lies in this interval (at the two extreme prices, we expect the profits to be small or negative). In the American put option problem (also discussed in §5), we have constraints that are handled explicitly. These are described in §5.

Constraint potentials share the same units as discrete potentials, and so it is important to not confuse the two. In an ID representation, discrete potentials are associated with discrete chance variables, and constraint potentials are associated with decision variables. We will sometimes refer to the set of discrete and continuous potentials as *probability potentials*, which does not include constraint potentials.

3.7. Utility Potentials

An ID representation includes utility functions that represent the preferences of the decision maker for the various outcomes. If an ID has more than one utility node, we assume an additive factorization of the joint utility function. Each additive factor of the utility function is represented by a utility potential. Formally, a *utility potential* v for $t \subseteq \mathcal{V}$ is a function $v: \Omega_t \rightarrow \mathbb{R}$ such that the values (in \mathbb{R}) are in units of utiles. An example of an utility potential is found in an example described below.

3.8. Summary

In summary, we can have four different kinds of potentials in IDs. The values of discrete potentials are in units of probabilities, which are dimensionless numbers (in the interval $[0, 1]$) with no physical units. The values of continuous potentials are in units of density, such as $(\text{unit}X)^{-1}$, $(\text{unit}X)^{-1} \cdot (\text{unit}Y)^{-1}$, etc. The values of utility potentials are in units of utiles. The values of constraint potentials are either

0 or 1 in dimensionless units. In the process of solving an ID, we may create potentials that have hybrid units such as $\text{utiles} \cdot (\text{unit}X)^{-1}$, etc. However, after we marginalize all chance and decision variables, we will ultimately end with a utility potential for the empty set. Details are provided in the section on solving hybrid IDs.

3.9. An Example

We will illustrate the concepts described so far using the entrepreneur's problem adapted from Howard (1971). An entrepreneur has to decide on a price for her new product. When the entrepreneur selects a price P (in dollars per widget), the quantity Q_n (in m widgets) that she will sell is determined from the demand curve $Q_n(P)$. This quantity Q_n will have a total cost of manufacturing $C_n(Q_n)$ (in m \$) given by the total cost curve. The entrepreneur's profit π_n (in m \$) will then be the difference between her revenue $P \cdot Q_n$ and her cost C_n , i.e., $\pi_n = P \cdot Q_n - C_n$. We assume that the entrepreneur is risk neutral, i.e., her utility is linear in millions of dollars, $u(xm\$) = x$ utiles. The entrepreneur needs to decide on a price p that will maximize her utility.

This problem would be simple if the demand curve and total cost curve were known with certainty, but this is seldom the case. We shall assume that the quantity Q_n determined from the demand curve is only a nominal value and that the actual quantity sold will be $Q_a = Q_n + Z_1$, where Z_1 (in m widgets) is a standard normal random variable. Furthermore, producing this quantity Q_a will cost $C_a = C_n(Q_a) + Z_2$, where Z_2 (in m \$) is another independent standard normal random variable. Note that the profit π (in m \$) is now $\pi = P \cdot Q_a - C_a$.

For the demand curve, the functional form is $Q_n(p) = (\ln \alpha - \ln p) / \beta$, where $p \leq \alpha$, and the constants are given by $\alpha = 50$ and $\beta = 1/80$. This is a decreasing function—at a price of \$1/widget, she would sell $80 \cdot \ln 50 \approx 313$ m widgets, and at a price of \$50/widget, she would sell none. For the total cost function we assume the form $C_n(q_a) = k_0 + k_1 q_a + k_2(1 - e^{-k_3 q_a})$ with constants $k_0 = 700$, $k_1 = 4$, $k_2 = 400$, and $k_3 = 1/50$. The total cost function is an increasing function, but at a decreasing rate. We restrict the range of P to make sure that Q_a is nonnegative. An ID representation of the problem is depicted in Figure 1.

The potentials in this example are as follows. We start with the name of the potential and then give its domain, details of the potential, and its units:

1. χ_a for $\{C_n, Z_2, C_a\}$ such that $\chi_a(c_n, z_2, c_a) = \delta(c_a - (c_n + z_2))$, $(m\$)^{-1}$;
2. φ_2 for Z_2 such that $\varphi_2(z_2) = (1/\sqrt{2\pi})e^{-z_2^2/2}$, $(m\$)^{-1}$;
3. χ_n for $\{Q_n, C_n\}$ such that $\chi_n(q_n, c_n) = \delta(c_n - (700 + 4q_n + 400(1 - e^{-q_n/50})))$, $(m\$)^{-1}$;
4. θ_a for $\{Q_n, Z_1, Q_a\}$ such that $\theta_a(q_n, z_1, q_a) = \delta(q_a - (q_n + z_1))$, $(mwidgets)^{-1}$;
5. φ_1 for Z_1 such that $\varphi_1(z_1) = (1/\sqrt{2\pi})e^{-z_1^2/2}$, $(mwidgets)^{-1}$;
6. θ_n for $\{P, Q_n\}$ such that $\theta_n(p, q_n) = \delta(q_n - 80(\ln 50 - \ln p))$, $(mwidgets)^{-1}$;
7. π for $\{P, Q_n, C_n\}$ such that $\pi(p, q_n, c_n) = p \cdot q_n - c_n$, utiles.

It is evident from the units that the first six potentials are continuous potentials and the seventh is a utility potential. There is no potential associated with decision variable P . A valuation network (VN) representation (Shenoy 1992) (also called a factor graph by Kschischang et al. 2001) of the entrepreneur's problem is shown in Figure 2. A VN is a bipartite graph with variables and potentials as nodes. Variable nodes are depicted just as in IDs. Potential nodes are depicted by diamond-shaped nodes. Probability and utility potentials are depicted by single-bordered diamond-shaped nodes. Constraint potentials are depicted by double-bordered diamond-shaped nodes. Each potential has an edge between it and the variables in its domain. During the solution phase, we switch from the ID to the VN representation because there are no guarantees that the ID representation will be maintained at each step of the solution algorithm (Shenoy 1992). In §5, we describe a solution to this problem.

3.10. Combination of Potentials

The definition of combination of potentials depends on the units of the potentials being combined. Although there are many possible combinations of units, we have only two distinct definitions. Utility functions are additive factors of the joint utility function. Thus, combination of two utility potentials (both in units of utiles) involves pointwise addition. In all other cases, combination of potentials involves pointwise multiplication. Thus, in problems where we have a single utility node, combination is always pointwise multiplication.

Suppose v_1 and v_2 are utility potentials for t_1 and t_2 , respectively. Then, the combination of v_1 and v_2 , denoted by $v_1 \otimes v_2$, is a utility potential for $t_1 \cup t_2$ defined as follows:

$$(v_1 \otimes v_2)(\mathbf{x}) = v_1(\mathbf{x}^{\downarrow t_1}) + v_2(\mathbf{x}^{\downarrow t_2}) \quad \text{for all } \mathbf{x} \in \Omega_{t_1 \cup t_2}. \quad (1)$$

The units of $(v_1 \otimes v_2)$ are utiles.

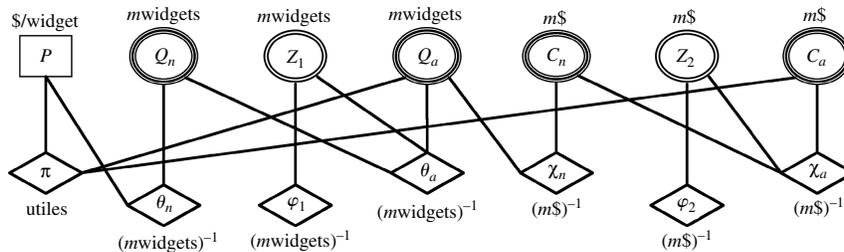
Suppose α_1 and α_2 are potentials for t_1 and t_2 , respectively, such that α_1 and α_2 are not both utility. Then, the combination of α_1 and α_2 , denoted by $\alpha_1 \otimes \alpha_2$, is a potential for $t_1 \cup t_2$ defined as follows:

$$(\alpha_1 \otimes \alpha_2)(\mathbf{x}) = \alpha_1(\mathbf{x}^{\downarrow t_1}) \cdot \alpha_2(\mathbf{x}^{\downarrow t_2}) \quad \text{for all } \mathbf{x} \in \Omega_{t_1 \cup t_2}. \quad (2)$$

The units of $(\alpha_1 \otimes \alpha_2)$ are the product of the units of α_1 and α_2 . Thus, e.g., if α_1 is discrete and α_2 is utility (or vice versa), then $\alpha_1 \otimes \alpha_2$ is utility; and if α_1 is continuous and α_2 is utility (or vice versa), then $\alpha_1 \otimes \alpha_2$ will have hybrid units such as utiles \cdot (unitX) $^{-1}$, etc.

Observe that combination of potentials is nonassociative. Thus, if σ is a discrete or continuous potential, and v_1 and v_2 are utility potentials, then $\sigma \otimes (v_1 \otimes v_2) \neq (\sigma \otimes v_1) \otimes v_2$. This nonassociativity of combination will necessitate divisions if we wish to use local

Figure 2 A Valuation Network Representation of the Entrepreneur's Problem



computation (Shenoy 1992). This will be discussed further in the section on solving hybrid IDs.

3.11. Marginalization of Potentials

In the process of solving an ID, we marginalize chance and decision variables in some sequence that is dictated by the information constraints. Before we marginalize a variable, we may have to do some combination and division operations prior to marginalization. In this subsection, we define just the marginalization operation without describing the details of how the potential being marginalized is obtained. The details of the solution algorithm are described after we have completed all requisite definitions.

The definition of marginalization of potentials depends on the nature of the variable being marginalized. We marginalize discrete chance variables by addition over its state space, continuous chance variables by integration over its state space, and decision variables (discrete or continuous) by maximization over its state space, which may be further constrained by constraint potentials.

Suppose α is a potential for a , and suppose $X \in a$ is a discrete variable. Then, the marginal of α by deleting X , denoted by α^{-X} , is a potential for $a \setminus \{X\}$ given as follows:

$$\alpha^{-X}(\mathbf{y}) = \sum_{x \in \Omega_X} \alpha(x, \mathbf{y}) \quad \text{for all } \mathbf{y} \in \Omega_{a \setminus \{X\}}. \quad (3)$$

In this case, the units of α^{-X} are exactly the same as the units of α .

If $X \in a$ is a continuous variable, then α^{-X} is defined as follows:

$$\alpha^{-X}(\mathbf{y}) = \int_{-\infty}^{\infty} \alpha(x, \mathbf{y}) dx \quad \text{for all } \mathbf{y} \in \Omega_{a \setminus \{X\}}. \quad (4)$$

In this case, the units of α^{-X} are the units of α multiplied by the units of X .

And if $X \in a$ is a decision variable, then α^{-X} is defined as follows:

$$\alpha^{-X}(\mathbf{y}) = \max_{x \in \Omega_X} \alpha(x, \mathbf{y}) \quad \text{for all } \mathbf{y} \in \Omega_{a \setminus \{X\}}. \quad (5)$$

In this case, the units of α^{-X} are exactly the same as the units of α . If we have a constraint potential χ for s associated with $X \in s$, then we assume that χ is already included in α (so that $s \subseteq a$), and the maximization in Equation (5) is over $x \in \Omega_X$ such that $\chi(x, \mathbf{y}^{\downarrow s \setminus \{X\}}) = 1$.

3.12. Division of Potentials

The process of solving an ID may involve division of discrete or continuous potentials by discrete or continuous potentials. Also, the potential in the divisor is always a marginal of the potential being divided.

Suppose α is a discrete or continuous potential for a , and suppose $X \in a$ is a discrete or continuous chance variable. Then the division of α by α^{-X} , denoted by $\alpha \oslash \alpha^{-X}$, is a potential for a defined as follows:

$$(\alpha \oslash \alpha^{-X})(x, \mathbf{y}) = \alpha(x, \mathbf{y}) / \alpha^{-X}(\mathbf{y}) \quad \text{for all } x \in \Omega_X, \text{ and } \mathbf{y} \in \Omega_{a \setminus \{X\}}. \quad (6)$$

In Equation (6), if the denominator is 0, then the numerator is also 0, and in this case we define 0/0 as 0. The units of the potential $\alpha \oslash \alpha^{-X}$ are the units of α divided by the units of α^{-X} . For the division operations that are done in the process of solving an ID (described in the next subsection) it can be shown that $\alpha \oslash \alpha^{-X}$ represents the conditional for X given variables in $a \setminus \{X\}$. Thus, if X is discrete, then $\alpha \oslash \alpha^{-X}$ is discrete, and if X is continuous, $\alpha \oslash \alpha^{-X}$ is continuous in units of $(\text{unit } X)^{-1}$ (Ciniciglu and Shenoy 2009).

3.13. An Algorithm for Solving Hybrid Influence Diagrams

We have all the definitions needed to solve hybrid IDs with deterministic conditionals. The solution algorithm is basically the same as described by Shenoy (1992) and Lauritzen and Nilsson (2001) for discrete IDs. The details of the solution algorithm are as follows.

First, all variables need to be marginalized in a sequence that respects the information constraints in the sense that if X precedes Y in the information sequence, then Y must be marginalized before X . In a well-defined ID, the information constraints form a partial order such that if C is a chance variable, and D is a decision variable, exactly one of the following information constraints must hold: either C precedes D , or D precedes C . In the former case, the true value of C is known by the decision maker prior to choosing a state of D , and in the latter case, the true value of C is not known at the time the decision maker has to choose a state of D .

First, we describe the general case where we have an additive factorization of the joint utility function. In

this case, divisions may be required. Next, we describe some special cases where divisions can be avoided.

We start with a set of potentials included in an ID representation. These potentials get modified in the process of marginalization.

3.14. Marginalizing a Chance Variable Case 1

Suppose we have to marginalize a chance variable C . First, we combine all probability potentials whose domains include C , resulting in the potential, say, χ . Next, we compute the marginal χ^{-C} . Then, we compute the quotient $(\chi \oslash \chi^{-C})$. The set of all probability potentials whose domains include C are replaced by the potentials χ^{-C} and $(\chi \oslash \chi^{-C})$. The units of $(\chi \oslash \chi^{-C})$ are units of probability if C is discrete, and $(\text{unit}C)^{-1}$ if C is continuous. The operations described so far are equivalent to the operations involved in arc reversal (Olmsted 1983). Next, we combine all utility potentials that include C in their domains, resulting in utility potential, say, v . The set of all utility potentials that include C in their domains is now replaced by the potential v . Next, we replace v and $(\chi \oslash \chi^{-C})$ by the potential $(v \otimes (\chi \oslash \chi^{-C}))^{-C}$, which must be a utility potential. This concludes the end of the process of marginalizing C . After marginalizing chance variable C , there will not be any potentials that include C in their domains.

3.15. Marginalizing a Decision Variable Case 1

Suppose we have to marginalize decision variable D . First we combine all utility potentials that include D in their domains, and then we combine the resulting utility potential with constraint potentials for D , if any, resulting in utility potential, say, v . Next, we marginalize D from v . All utility and constraint potentials that include D in their domains are now replaced by v^{-D} . In the process of marginalizing D from v , we keep track of where the maximum is attained (as a function of the remaining variables in the domain of v). This yields a *decision function* for the decision variable. The collection of all decision functions constitutes an *optimal strategy* for the ID.

After all variables have been marginalized, we end up with a single utility potential for the empty set, whose value represents the optimal utility associated with an optimal strategy.

This general algorithm described above involves divisions in the process of marginalizing a chance

variable. This step may be simplified in the case where we have a single utility potential as follows.

3.16. Marginalizing a Chance/Decision Variable Case 2

Suppose we have to marginalize a (chance or decision) variable X . First, we combine all potentials that include X in their domains, resulting in potential, say, v , and then marginalize X from v . The set of all potentials that include X in their domains is replaced by v^{-X} . In this case, we cannot predict the nature of v^{-X} , i.e., it may have hybrid units.

Notice that there are no divisions involved in this process. When we have a single utility factor, combination always involves multiplication, which is associative, and it follows from the axiomatic approach of Shenoy and Shafer (1990) that we can find marginals without doing any divisions. The process of solving an ID can be described as finding the marginal for the empty set by sequentially marginalizing all variables in a sequence that respects the information constraints. The first example (entrepreneur's problem) solved in §5 has a single utility factor, and thus, no divisions are required.

Another special case where no divisions are necessary is as follows. In the process of marginalizing a chance variable C , suppose that there is only one probability (discrete or continuous) potential, say, χ , that includes C in its domain. In this case, χ must be the conditional for C given its parents, $pa(C)$. Thus, χ^{-C} is an identity potential for $pa(C)$ (whose values are all 1s). In this case also, we can skip the divisions. If this is true for all chance variables C (this will happen if the arcs into each chance variable are consistent with the partial order representing the information constraints and we pick a deletion sequence consistent with all arcs in the ID), then we can use the rules described in Case 2 above. The second example (American put option problem) solved in §5 is an example of this type, and no divisions are required.

Finally, we remark that one can always avoid divisions by combining all utility potentials and replacing the set of all utility potentials by the combination. This may, however, increase the computational effort of solving an ID because the domain of the single joint utility potential will have all decision variables in its domain and could potentially be

large. Shenoy (1992) described a small example where divisions are inescapable assuming that we wish to use local computation and avoid computing on the domain of all variables.

The algorithm described in this subsection is illustrated in §5 by solving two small hybrid IDs in complete detail.

4. Mixture of Polynomials Functions

In this section, we define MOP functions and describe some methods for finding MOP approximations of univariate and two-dimensional conditional PDFs and piecewise-linear approximations of nonlinear deterministic functions. We illustrate our method for the log-normal distribution. Shenoy and West (2011b) describe MOP approximations of the PDFs of the normal and chi-square univariate distributions, and CLG distributions in two dimensions. Shenoy (2012) describes MOP approximations of the CLG PDFs in one, two, and three dimensions.

4.1. MOP Functions

The definitions of one-dimensional and multidimensional MOP functions are taken from Shenoy (2012).

A one-dimensional function $f: \mathbb{R} \rightarrow \mathbb{R}$ is said to be a MOP function if it is a piecewise function of the form

$$f(x) = \begin{cases} a_{0i} + a_{1i}x + \dots + a_{ni}x^n & \text{for } x \in A_i, i = 1, \dots, k, \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where A_1, \dots, A_k are disjoint intervals in \mathbb{R} that do not depend on x , and a_{0i}, \dots, a_{ni} are constants for all i . We will say that f is a k -piece (ignoring the 0 piece) and n -degree (assuming $a_{ni} \neq 0$ for some i) MOP function.

An example of a two-piece, three-degree MOP function $g_1(\cdot)$ in one dimension is as follows:

$$g_1(x) = \begin{cases} 0.41035 + 0.09499x - 0.09786x^2 - 0.02850x^3 & \text{if } -3 < x < 0, \\ 0.41035 - 0.09499x - 0.09786x^2 + 0.02850x^3 & \text{if } 0 \leq x < 3, \\ 0 & \text{otherwise;} \end{cases} \quad (8)$$

$g_1(\cdot)$ is a MOP approximation of the PDF of the standard normal distribution on the domain $(-3, 3)$ and was found using Lagrange interpolating polynomials with Chebyshev points, which will be discussed in the next subsection.

The main motivation for defining MOP functions is that such functions are easy to integrate in closed form, and the family of MOP functions is closed under multiplication, addition, integration, the main operations in solving hybrid IDs. Also, because MOP functions are easily differentiable, it is easy to maximize MOP functions in closed form.

A multivariate polynomial is a polynomial in several variables. For example, a polynomial in two variables is as follows:

$$P(x_1, x_2) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 + a_{20}x_1^2 + a_{02}x_2^2 + a_{21}x_1^2x_2 + a_{12}x_1x_2^2 + a_{22}x_1^2x_2^2. \quad (9)$$

The degree of the polynomial in Equation (9) is 4 assuming a_{22} is a nonzero constant. In general, the degree of a multivariate polynomial is the largest sum of the exponents of the variables in the terms of the polynomial.

An m -dimensional function $f: \mathbb{R}^m \rightarrow \mathbb{R}$ is said to be a MOP function if

$$f(x_1, x_2, \dots, x_m) = \begin{cases} P_i(x_1, x_2, \dots, x_m) & \text{for } (x_1, x_2, \dots, x_m) \in A_i, \\ & i = 1, \dots, k, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where $P_i(x_1, x_2, \dots, x_m)$ are multivariate polynomials in m variables for all i , and the disjoint regions A_i are as follows. Suppose π is a permutation of $\{1, \dots, m\}$. Then, each A_i is of the form

$$\begin{aligned} l_{1i} &\leq x_{\pi(1)} \leq u_{1i}, \\ l_{2i}(x_{\pi(1)}) &\leq x_{\pi(2)} \leq u_{2i}(x_{\pi(1)}), \\ &\vdots \\ l_{mi}(x_{\pi(1)}, \dots, x_{\pi(m-1)}) &\leq x_{\pi(m)} \leq u_{mi}(x_{\pi(1)}, \dots, x_{\pi(m-1)}), \end{aligned} \quad (11)$$

where l_{1i} and u_{1i} are constants, and

$$l_{ji}(x_{\pi(1)}, \dots, x_{\pi(j-1)}) \text{ and } u_{ji}(x_{\pi(1)}, \dots, x_{\pi(j-1)})$$

are linear functions of $x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(j-1)}$ for $j = 2, \dots, m$ and $i = 1, \dots, k$. We will refer to the nature of the region described in Equation (11) as a *hyper-rhombus*. Although we have defined the hyper-rhombus as a closed region in Equation (11), each of the $2m$ inequalities can be either strictly $<$ or \leq . Notice that the definition of the region A_i in the m -dimensional case (in Equation (11)) is a generalization of the requirement in the one-dimensional case (Equation (7)) that the regions A_i are intervals.

A special case of the hyper-rhombus region A_i is a region of the form

$$l_{1i} \leq x_1 \leq u_{1i}, l_{2i} \leq x_2 \leq u_{2i}, \dots, l_{mi} \leq x_m \leq u_{mi}, \quad (12)$$

where l_{1i}, \dots, l_{mi} and u_{1i}, \dots, u_{mi} are all constants. We refer to the region defined in Equation (12) as a *hypercube* (in m dimensions).

An example of a two-piece, three-degree MOP $g_2(\cdot, \cdot)$ defined on a two-dimensional hyper-rhombus is as follows:

$$g_2(x, y) = \begin{cases} 0.41035 + 0.09499(y - x) - 0.09786(y - x)^2 \\ \quad - 0.02850(y - x)^3 & \text{if } x - 3 < y < x, \\ 0.41035 - 0.09499(y - x) - 0.09786(y - x)^2 \\ \quad + 0.02850(y - x)^3 & \text{if } x \leq y < x + 3, \\ 0 & \text{otherwise;} \end{cases} \quad (13)$$

$g_2(x, y)$ is a two-dimensional MOP approximation of the PDF of the CLG distribution of $Y | x \sim N(x, 1^2)$ on the domain $-\infty < x < \infty, x - 3 < y < x + 3$. Notice that $g_2(x, y) = g_1(y - x)$, where $g_1(\cdot)$ is as defined in Equation (8).

4.1.1. Advantages of Hyper-Rhombus Regions.

One advantage of defining multidimensional MOP functions on hyper-rhombuses is that MOP functions are closed under transformations needed for multidimensional linear deterministic conditionals. For example, consider the case where X, Y , and Z are continuous variables, where X has PDF $f_X(x)$, $Y | x$ has conditional PDF $f_{Y|x}(y)$, and Z has a deterministic conditional $Z = X + Y$, which is represented by the function $\delta(z - x - y)$, where δ is the

Dirac delta function. Suppose that $f_X(x)$ is a one-dimensional MOP function, and suppose that $f_{Y|x}(y)$ is a two-dimensional MOP function (in x and y) defined on hypercubes. Suppose we wish to find the marginal of Z . After we marginalize Y (by computing $\int_{-\infty}^{\infty} f_{Y|x}(y) \delta(z - x - y) dy$), we obtain the function $f_{Y|x}(z - x)$. Notice that even though $f_{Y|x}(y)$ was defined on hypercubes, $f_{Y|x}(z - x)$ is no longer defined on hypercubes because we now have regions such as $l_{1i} \leq z - x \leq u_{1i}$, which is a hyper-rhombus.

Another advantage is that we can obtain MOP approximations of CLG PDFs from a MOP approximation of the univariate standard normal PDF (Shenoy 2012). For example, suppose $g_1(x)$ is a MOP approximation of the PDF of $N(0, 1^2)$. Now suppose $Y | x \sim N(ax + b, \sigma^2)$, where a, b , and σ are constants, and $\sigma \neq 0$. We can find a MOP approximation of the PDF of $Y | x$ as follows:

$$h(x, y) = \frac{1}{|\sigma|} g_1\left(\frac{y - ax - b}{\sigma}\right). \quad (14)$$

Notice that even though $g_1(x)$ is defined on hypercubes, $h(x, y)$ is no longer defined on hypercubes (because we now have regions such as $l_{1i} \leq (y - ax - b)/\sigma \leq u_{1i}$). However, $h(x, y)$ is defined on a hyper-rhombus region, and therefore is a MOP. The MOP function $g_2(x, y)$ described in Equation (13) is an instance of $h(x, y)$ when $a = 1, b = 0$, and $\sigma = 1$.

Finally, the hyper-rhombus region allows us to find MOP approximations of conditional PDFs using fewer pieces and lower degrees. Using hypercubes, we were able to find a 16-piece, 18-degree MOP approximation of a conditional log-normal PDF. Using hyper-rhombuses, we found an eight-piece, five-degree MOP approximation for the same conditional log-normal PDF. This is because in the hyper-rhombus case, we can truncate the region where the PDF has very small values, and thus avoid the high degree necessitated by the nonnegativity condition of PDFs.

There are some disadvantages associated with hyper-rhombus regions compared to hypercubes. MOPs defined on hyper-rhombuses take longer to integrate. After integration, MOPs defined on hyper-rhombuses may have higher degrees. Some comparisons of hyper-rhombuses versus hypercubes appear in Shenoy (2012) and Shenoy et al. (2011).

4.2. Finding MOP Approximations of Univariate PDFs

In this subsection, we will describe a process for finding a MOP approximation of a univariate PDF using Lagrange interpolating polynomials with Chebyshev points. In the next section, we will work with log-normal PDFs. Therefore, we will use the log-normal distribution for illustration purposes.

4.2.1. Lagrange Interpolating Polynomials. Suppose we need to fit a polynomial for a one-dimensional function $f(x)$ in some interval (a, b) . Given a set of n points $\{(x_1, f(x_1)), \dots, (x_n, f(x_n))\}$, the Lagrange interpolating polynomial (LIP) $P(x)$ is given by

$$P(x) = \sum_{j=1}^n \left[f(x_j) \prod_{k=1, k \neq j}^n \frac{x - x_k}{x_j - x_k} \right]. \quad (15)$$

The polynomial $P(x)$ has the following properties (Burden and Faires 2010). It is a polynomial of degree $\leq (n - 1)$ that passes through the n points $\{(x_1, f(x_1)), \dots, (x_n, f(x_n))\}$, i.e., $P(x_j) = f(x_j)$ for $j = 1, \dots, n$. If $f(x)$ is continuous and $(n + 1)$ -times differentiable in an interval (a, b) , and x_1, \dots, x_n are distinct points in (a, b) such that $x_1 < \dots < x_n$, then for each $x \in (a, b)$, there exists a number $\xi(x)$ (generally unknown) between x_1 and x_n such that

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi(x))}{n!} (x - x_1)(x - x_2) \cdots (x - x_n). \quad (16)$$

4.2.2. Chebyshev Points. One question in the use of LIP is the choice of the points x_1, \dots, x_n . For an interval (a, b) , where $b > a$, the n Chebyshev points are given by

$$x_i = \frac{1}{2}(a + b) + \frac{1}{2}(b - a) \cos\left(\frac{2i - 1}{2n} \pi\right), \quad \text{for } i = 1, \dots, n. \quad (17)$$

The Chebyshev points are often used with LIP because the resulting polynomial approximation $P(x)$ minimizes the quantity $|(x - x_1) \cdots (x - x_n)|$ for all $x \in (a, b)$, which is proportional to the absolute error between the function $f(x)$ and the LIP $P(x)$ (see Equation (16)). The minimum value of $|(x - x_1) \cdots (x - x_n)|$ is $1/2^{n-1}$. Thus, as n increases, the maximum absolute deviation decreases.

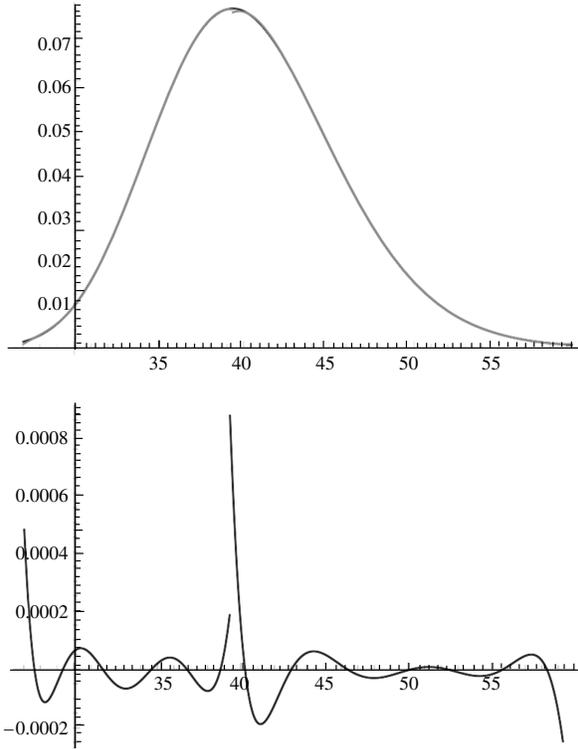
4.2.3. Finding a MOP Approximation for a PDF.

The construction of a k -piece, n -degree MOP $g(x)$ that approximates a PDF $f(x)$ on some domain (l, u) proceeds as follows. First, we compute the LIP polynomial, say, $g_u(x)$, for $f(x)$ using $n = 3$ Chebyshev points for the domain (l, u) . Second, we check to see if $g_u(x)$ is nonnegative over the entire domain (by computing the minimum of $g_u(x)$ over the entire domain and making sure it is positive). If not, we increase n until we obtain nonnegativity. Because we are using Chebyshev points, we are guaranteed to obtain nonnegativity for some n assuming $f(x) > 0$ for $x \in (l, u)$. If the smallest degree n for which we obtain nonnegativity is too high (> 5 , e.g., for a one-dimensional MOP), then we partition the domain into more pieces and restart. Third, we normalize the fitted polynomial $g_u(x)$ so that it integrates to 1.

The procedure described in the previous paragraph can be applied to any PDF, including, e.g., the class of quantile-parameterized distributions described by Keelin and Powley (2011). We will apply the above procedure for a log-normal distribution. Suppose $X \sim N(\mu, \sigma^2)$ and $Y = e^X$. Then, we say, Y has the log-normal distribution with parameters μ and σ^2 , written as $Y \sim LN(\mu, \sigma^2)$. First, we need to decide on the precision of the MOP approximation. The exact domain of the PDF of Y is $(0, \infty)$. For the standard normal distribution, the domain $(-3, 3)$ covers 99.7% of the total probability. Thus, we can approximate the PDF of Y on the domain $(e^{\mu-3\sigma}, e^{\mu+3\sigma})$. If we need greater precision, we can approximate the PDF of Y on a larger domain, e.g., on $(e^{\mu-4\sigma}, e^{\mu+4\sigma})$, which captures more than 99.99% of the total probability.

Suppose $S_1 \sim LN(\mu, \sigma^2)$, where $\mu = \ln(40) + 0.00074$ and $\sigma^2 = 0.13229^2$ (these constants are obtained from the American put option example described in §5.2). The 0.15 percentile of the PDF of S_1 is 27.03, and the 99.85 percentile is 59.28. If we try to fit a one-piece MOP approximation of the PDF of S_1 using the above procedure, the result is an eight-degree MOP on the domain (27.03, 59.28). So we partition the domain into two pieces, (27.03, 39.34) and [39.34, 59.28), where 39.34 is the mode of S_1 ($= e^{\mu - \sigma^2}$). In this case, we find a two-piece, five-degree MOP $\phi_{p1}(s_1)$. A graph of the MOP approximation $\phi_{p1}(s_1)$ overlaid on the actual PDF $\phi_1(s_1)$ truncated to

Figure 3 A Graph of ϕ_{p1} (in Grey) Overlaid on ϕ_1 (in Dark Grey); the Bottom Graph Shows the Difference $\phi_1 - \phi_{p1}$ in the Curves



(27.03, 59.28) is shown in Figure 3, and it shows that there are not much differences between the two functions. The mean of $\phi_{p1}(s_1)$ is ≈ 40.370 , and the mean of $\phi_1(s_1) \approx 40.371$. The variance of $\phi_{p1}(s_1)$ is ≈ 27.875 , and the variance of $\phi_1(s_1) \approx 27.879$.

The LIP method described for univariate PDFs can also be applied for two-dimensional conditional PDFs. A generalization of Lagrange interpolating polynomials exists for two and higher dimensions and its implementations exist in commercial software such as Mathematica and Maple. Chebyshev points are also defined in closed form for two-dimensional regions by Xu (1996).

4.3. Finding Piecewise-Linear Approximations of Deterministic Conditionals

In this subsection, we describe finding piecewise-linear approximations of deterministic conditionals. MOP functions are closed under transformation

required for multidimensional linear and quotient functions. Thus, if we have a nonlinear deterministic conditional (as we do in the entrepreneur's problem), then we need to approximate such conditionals by piecewise linear functions.

Consider the nominal demand function f_{Q_n} as a function of price p in the range $1 \leq p \leq 47$ given as follows:

$$f_{Q_n}(p) = 80(\ln 50 - \ln p) \quad \text{if } 1 \leq p \leq 47. \quad (18)$$

Because $f_{Q_n}(p)$ is not a linear function, the deterministic conditional $Q_n = f_{Q_n}(P)$ (associated with Q_n) is not linear. Thus, we need to approximate $f_{Q_n}(p)$ by a piecewise linear function. We select a set of intermediate points in the interval $(1, 47)$ and find a piecewise linear approximation $f_{pQ_n}(p)$ of this one-dimensional function as follows:

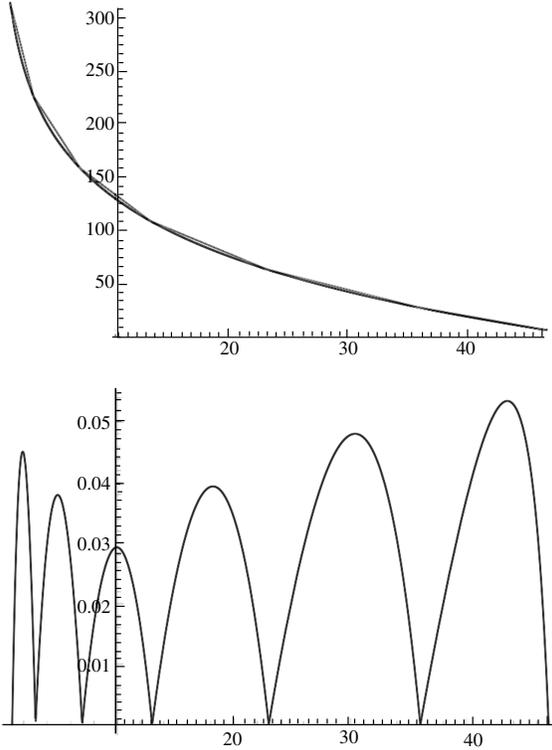
$$f_{pQ_n}(p) = \begin{cases} 225.073 - 43.9445(p - 3) & \text{if } 1 \leq p < 3, \\ 157.289 - 16.946(p - 7) & \text{if } 3 \leq p < 7, \\ 107.766 - 8.25386(p - 13) & \text{if } 7 \leq p < 13, \\ 69.4 - 4.79573(p - 21) & \text{if } 13 \leq p < 21, \\ 28.534 - 2.919(p - 35) & \text{if } 21 \leq p < 35, \\ 4.95003 - 1.96533(p - 47) & \text{if } 35 \leq p \leq 47, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

The number of intermediate points and the location of the points were selected by trial and error. A graph of $f_{pQ_n}(p)$ versus p overlaid on the graph of $f_{Q_n}(p)$ versus p is shown in Figure 4. The maximum absolute percentage deviation between $f_{Q_n}(p)$ and $f_{pQ_n}(p)$ is 5.3% at $p = 43.5$.

5. Two Examples

In this section, we illustrate our framework and algorithm for solving hybrid IDs with deterministic conditionals by solving two problems. The first one is the entrepreneur's problem described in §3, and has continuous chance and deterministic conditionals, a continuous decision variable, and one (unfactored) utility function. The second problem is an American put option described by Charnes and Shenoy (2004). This problem has continuous chance variables, discrete decision variables with continuous

Figure 4 A Graph of $f_{pQ_n}(p)$ (in Grey) Overlaid on $f_{Q_n}(p)$ (in Dark Grey); the Bottom Graph Shows the Difference $f_{pQ_n} - f_{Q_n}$ in the Curves



chance predecessors, and an additive factorization of the utility function. For some of the complicated marginalization operations, we report the approximate time it takes Mathematica to do the operation (using the *Timing* command in Mathematica). We used Mathematica version 7.0.1 on a MacBook Pro laptop computer to do the computations.

5.1. Entrepreneur's Problem.

We will solve the entrepreneur's problem by marginalizing variables in the following sequence: C_a , Z_2 , C_n , Q_a , Z_1 , Q_n , P . To avoid the integration and optimization problems, we will approximate the continuous potentials associated with Z_1 and Z_2 by MOP potentials φ_{p1} and φ_{p2} , respectively, and the nonlinear deterministic functions associated with Q_n and C_n by piecewise-linear functions f_{pQ_n} and f_{pC_n} , respectively. Because we have a single utility potential, no divisions are necessary during the solution process.

5.1.1. Marginalizing C_a . First, we marginalize C_a ; C_a is in the domain of potentials χ_a and π . Let potential π_1 denote $(\chi_a \otimes \pi)^{-C_a}$. Then,

$$\begin{aligned} \pi_1(p, q_a, c_n, z_2) &= (\chi_a \otimes \pi)^{-C_a}(p, q_a, c_n, z_2) \\ &= \int_{-\infty}^{\infty} \delta(c_a - (c_n + z_2)) \cdot (p \cdot q_a - c_a) \, dc_a \\ &= p \cdot q_a - (c_n + z_2) \quad (\text{utils}). \end{aligned} \quad (20)$$

The result in Equation (20) follows from Property 1 of Dirac delta functions.

5.1.2. Marginalizing Z_2 . Next, we marginalize Z_2 ; Z_2 is in the domain of potentials φ_{p2} and π_1 . Let $\varphi_{p2}(z)$ denote the two-piece, three-degree MOP approximation of $\varphi_2(z)$, the PDF associated with the standard normal distribution, as described in Equation (8).

Let π_2 denote $(\varphi_{p2} \otimes \pi_1)^{-Z_2}$. Details of π_2 are as follows:

$$\begin{aligned} \pi_2(p, q_a, c_n) &= (\varphi_{p2} \otimes \pi_1)^{-Z_2}(p, q_a, c_n) \\ &= \int_{-\infty}^{\infty} \varphi_{p2}(z_2) \cdot (p \cdot q_a - (c_n + z_2)) \, dz_2 \\ &= p \cdot q_a - c_n \quad (\text{utils}). \end{aligned} \quad (21)$$

5.1.3. Marginalizing C_n . Next we marginalize C_n ; C_n is in the domain of potentials χ_n and π_2 . Let f_{pC_n} denote a three-piece piecewise-linear approximation of the cost function f_{C_n} as follows:

$$f_{pC_n}(q_a) = \begin{cases} 705.1 + 9.29q_a & \text{if } 2 \leq q_a < 42, \\ 844.19 + 5.98q_a & \text{if } 42 \leq q_a < 104, \\ 1,025.87 + 4.23q_a & \text{if } 104 \leq q_a \leq 316, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

The maximum absolute percentage deviation between f_{pC_n} and f_{C_n} is 2.3% at $q_a = 18.64$. The Dirac potential associated with C_n is $\chi_n(c_n, q_a) = \delta(c_n - f_{pC_n}(q_a))$. Let π_3 denote $(\chi_n \otimes \pi_2)^{-C_n}$. Details of π_3 are as follows:

$$\begin{aligned} \pi_3(p, q_a) &= (\chi_n \otimes \pi_2)^{-C_n}(p, q_a) \\ &= \int_{-\infty}^{\infty} (pq_a - c_n) \cdot \delta(c_n - f_{pC_n}(q_a)) \, dc_n \\ &= pq_a - f_{pC_n}(q_a) \quad (\text{utils}). \end{aligned} \quad (23)$$

5.1.4. Marginalizing Q_a . Next, we marginalize Q_a ; Q_a is in the domain of potentials θ_a and π_3 . Let π_4 denote $(\theta_a \otimes \pi_3)^{-Q_a}$; details of π_4 are as follows:

$$\begin{aligned}\pi_4(p, q_n, z_1) &= (\theta_a \otimes \pi_3)^{-Q_a}(p, q_n, z_1) \\ &= \int_{-\infty}^{\infty} \delta(q_a - (q_n + z_1)) \cdot (p \cdot q_a - f_{pC_n}(q_a)) dq_a \\ &= p \cdot (q_n + z_1) - f_{pC_n}(q_n + z_1) \quad (\text{utils}). \quad (24)\end{aligned}$$

Notice that π_4 is a MOP function.

5.1.5. Marginalizing Z_1 . Next, we marginalize Z_1 , which is in the domain of potentials φ_{p1} and π_4 . Let π_5 denote $(\varphi_{p1} \otimes \pi_4)^{-Z_1}$; π_5 is computed as follows:

$$\begin{aligned}\pi_5(p, q_n) &= (\varphi_{p1} \otimes \pi_4)^{-Z_1}(p, q_n) \\ &= \int_{-\infty}^{\infty} \varphi_{p1}(z_1) \cdot (p(q_n + z_1) - f_{pC_n}(q_n + z_1)) dz_1 \\ &= p \cdot q_n - \int_{-\infty}^{\infty} f_{pC_n}(q_n + z_1) \cdot \varphi_{p1}(z_1) dz_1. \quad (25)\end{aligned}$$

Notice that because φ_{p1} and f_{pC_n} are MOP functions, π_5 is also a MOP function (15 pieces, 5 degree). It takes Mathematica about 3.7 seconds to do the integration in Equation (25).

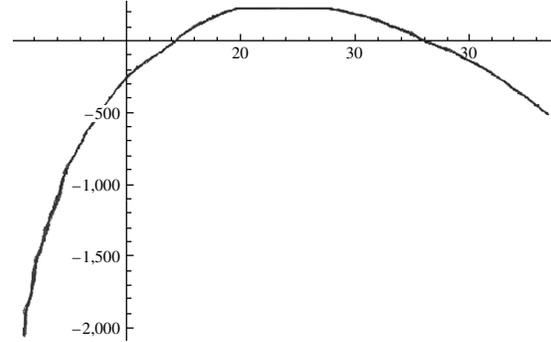
5.1.6. Marginalizing Q_n . Next, we marginalize Q_n ; Q_n is in the domain of potentials θ_n and π_5 . Let π_6 denote $(\theta_n \otimes \pi_5)^{-Q_n}$. The details of π_6 are as follows:

$$\begin{aligned}\pi_6(p) &= (\theta_n \otimes \pi_5)^{-Q_n}(p) \\ &= \int_{-\infty}^{\infty} \delta(q_n - f_{pQ_n}(p)) \cdot \pi_5(p, q_n) dq_n \\ &= \pi_5(p, f_{pQ_n}(p)). \quad (26)\end{aligned}$$

Because $f_{pQ_n}(p)$ is a piecewise-linear function, and π_5 is a MOP function, $\pi_6(p)$ is a MOP function; π_6 is computed as a 15-piece, 5-degree MOP function. It takes Mathematica about 14.8 seconds to do the integration in Equation (26).

5.1.7. Marginalizing P . Figure 5 shows a graph of $\pi_6(p)$ versus p . Finally, we marginalize P . The maximum utility is 234.12 utils at $p = \$25.76$ per widget. It takes Mathematica 0.15 seconds to marginalize P from π_6 . For comparison, when demand and supply are known with certainty, the problem reduces to a nonlinear optimization problem and the maximum utility 198 utils is obtained when price is \$24.10 per widget.

Figure 5 A Graph of $\pi_6(p)$ vs. p

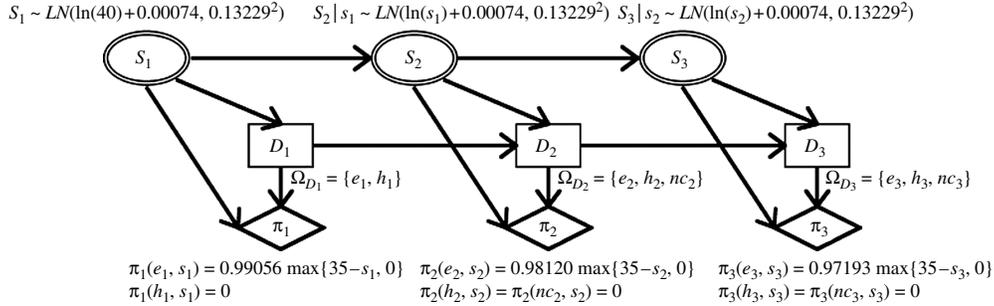


5.2. An American Put Option Problem

This problem is adapted from Charnes and Shenoy (2004). An option trader has to decide whether or not to exercise a seven-month put option with initial stock price $S_0 = \$40$ and exercise price $X = \$35$. A put option on a stock provides the owner of the option the right to sell one share of the stock at the exercise price during the period of the option. For example, if the price of the stock dips to, say, \$30, during the option period, then the owner of the option described above can buy one share at \$30 and sell it for \$35, with a realized profit of \$5. In reality, the option can be exercised at any time before the expiration of the option. For modeling purposes, we assume that the option is available for exercise at three equally spaced decision points over a seven-month period. Following standard practice in the financial literature, the stock prices, S_1, S_2, \dots, S_k , evolve according to the discrete stochastic process: $S_j = S_{j-1} \cdot Y$, where $Y \sim LN((r - \sigma^2/2)\Delta t, \sigma^2\Delta t)$, for $j = 1, 2, \dots, k$, S_j is the stock price (in dollars) at time $j\Delta t$, r is the risk-less interest rate (per year), σ is the stock's volatility (per year), T denotes the length of the option (in years), and $\Delta t = T/k$. We assume $r = 0.0488$ per year, $T = 0.5833$ years, $\Delta t = 0.1944$ years, $k = 3$ stages, and $\sigma = 0.3$ per year (these constants are borrowed from Geske and Johnson 1984, which provides an analytic value of the option for comparison purposes). Thus, $S_1 \sim LN(\ln 40 + 0.00074, 0.13229^2)$, $S_2 | s_1 \sim LN(\ln s_1 + 0.00074, 0.13229^2)$, and $S_3 | s_2 \sim LN(\ln s_2 + 0.00074, 0.13229^2)$. An ID representation of the problem is shown in Figure 6.

The state space of D_1 is $\{e_1, h_1\}$, i.e., exercise or hold. The constraints for the decision nodes D_2 and D_3

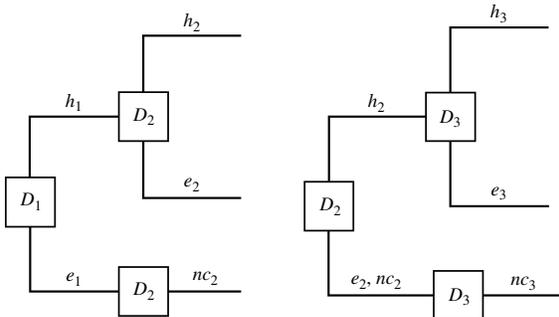
Figure 6 An Influence Diagram Representation of the American Put Option Problem



in the problem are shown in Figure 7, where e_i, h_i , and nc_i denote the alternatives: exercise, hold, and no choice, respectively, for decision $D_i, i = 2, 3$. The only possible decision for stage i is no choice if the stock was exercised at a prior time. The additive factors of the utility function are $\pi_j(d_j, s_j) = e^{-rj\Delta t} \max\{35 - s_j, 0\}$, if $d_j = e_j$, and $\pi_j = 0$ otherwise. The $e^{-rj\Delta t}$ is a discount factor to translate future profits back to the present ($j = 0$). As in the entrepreneur’s problem, we assume that the decision maker’s utilities for profits are linear in dollars.

We approximate the marginal PDF of S_1 by a MOP function $\phi_{p1}(s_1)$. Also, the MOP approximations of the conditional PDFs for $S_2 | s_1$ and $S_3 | s_2$ are denoted by $\psi_{p2}(s_1, s_2)$ and $\psi_{p3}(s_2, s_3)$, respectively. Also, we model the constraints on the choices at D_2 and D_3 by constraint potentials χ_2 for D_2 in $\{D_1, D_2\}$ and χ_3 for D_3 in $\{D_2, D_3\}$. The values of χ_2 and χ_3 are 1s for the possible states (as shown in Figure 7) and 0s for the rest.

Figure 7 Constraints on D_2 and D_3



The potentials in the problem are as follows (name, domain, and units):

π_3 ,	for $\{S_3, D_3\}$,	utils;
χ_3 ,	for $D_3 \in \{D_2, D_3\}$,	no units;
ψ_{p3} ,	for $\{S_2, S_3\}$,	$(\$)^{-1}$;
χ_2 ,	for $D_2 \in \{D_1, D_2\}$,	no units;
π_2 ,	for $\{S_2, D_2\}$,	utils;
ψ_{p2} ,	for $\{S_1, S_2\}$,	$(\$)^{-1}$;
π_1 ,	for $\{S_1, D_1\}$,	utils;
ϕ_{p1} ,	for $\{S_1\}$,	$(\$)^{-1}$.

The information constraints in the ID constrain us to marginalize the variables in the following sequence: $D_3, S_3, D_2, S_2, D_1, S_1$. Because the conditional arcs for S_1, S_2 , and S_3 are consistent with the partial order determined by the information constraints, no divisions are required.

5.2.1. Marginalizing D_3 and S_3 First we marginalize D_3 . Because D_3 is in the domains of potentials π_3 and χ_3 , we first combine these and then marginalize D_3 from the combination. Let π'_3 denote $(\chi_3 \otimes \pi_3)^{-D_3}$. The units of values of π'_3 are utils. Because $\pi_3(e_3, s_3) \geq \pi_3(h_3, s_3)$ and $\pi_3(e_3, s_3) \geq \pi_3(nc_3, s_3)$ for all values of $s_3 < 35$, the details of π'_3 are as follows:

$$\pi'_3(d_2, s_3) = \begin{cases} 0.97 \max\{35 - s_3, 0\} & \text{if } d_2 = h_2, \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Thus, the optimal strategy of stage 3 would be to exercise the option if the observed value of $S_3 < 35$, assuming this alternative is available (i.e., the option has not been exercised earlier), and to abandon the option if the observed value of $S_3 \geq 35$.

Next, we marginalize S_3 , which is in the domains of ψ_{p_3} and π'_3 . Let π''_3 denote $(\psi_{p_3} \otimes \pi'_3)^{-S_3}$. The units of values of π''_3 are in utiles. Details of π''_3 are as follows:

$$\pi''_3(d_2, s_2) = \begin{cases} \int_0^{35} \pi'_3(d_2, s_3) \cdot \psi_{p_3}(s_2, s_3) ds_3 & \text{if } d_2 = h_2, \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

Because π'_3 and ψ_{p_3} are MOP functions, π''_3 is a MOP function (four piece, nine degree).

Similarly, we marginalize the remaining variables. The optimal decision function at stage 2 is to exercise the option if the observed stock price is less than \$24.75 (assuming this option is available) or otherwise hold it for the next stage. The optimal decision function in stage 1 is to exercise the option when the stock price is less than \$28.15 or otherwise hold it for the next stage. The optimal value of the option is computed as \$1.219.

Our result is comparable to the financial analytic result \$1.219 (using the Black and Scholes (1973) option pricing theory computed analytically by Geske and Johnson (1984)), and the result \$1.224 computed by Monte Carlo method using 30 stages (Charnes and Shenoy 2004).

One practical benefit of solving this ID exactly is that not only do we get the value of the option (which is the focus of option pricing theory), we also get a strategy for exercising the option. The financial analytic result provides only the value of the option. The Monte Carlo method proposed by Charnes and Shenoy (2004) provides an approximate strategy by providing bounds on when to exercise the option. Our technique provides an exact strategy for the ID in which the conditional PDFs are approximated by MOP functions.

6. Summary and Conclusions

The main contribution of this paper is a framework and an algorithm for solving hybrid IDs with discrete and continuous chance variables, discrete and continuous decision variables, and deterministic conditionals for continuous chance variables.

First, the extended Shenoy-Shafer architecture for making inferences in hybrid BNs proposed by Shenoy and West (2011a) has been further extended to include

decision variables and utility functions. Second, we propose approximating conditional PDFs and utility functions by MOPs, and approximating nonlinear deterministic functions for continuous chance variables by piecewise linear functions. We have illustrated our framework and algorithm by solving two small hybrid IDs.

Two main problems in solving hybrid IDs are marginalization of continuous chance variables and marginalization of continuous decision variables. For decision problems that can be solved without divisions, one solution is to approximate conditional PDFs and utility functions by MOP functions, and nonlinear deterministic conditionals by piecewise linear functions. MOP functions are closed under multiplication, addition, integration, and under transformations needed for linear deterministic conditionals. However, they are not closed under divisions. Thus, MOP approximations could be used to mitigate the problems associated with marginalization of continuous chance and continuous decision variables when no divisions are needed. Also, it is relatively easier to maximize a utility function that is expressed in MOP form with a low degree. By solving for all real roots of a low-degree polynomial in closed form, we can compute a global maximum of the utility function as a function of other continuous variables in closed form.

There are two classes of decision problems that can be solved using local computation without doing any divisions. First, if we have a single utility function (with no additive factors), then combination is always multiplication, which is associative, and the axioms for local computations (see Shenoy and Shafer 1990) are satisfied by the combination and marginalization operations without needing any divisions. The entrepreneur's problem is an example of this genre. Second, if we have a decision problem where the arcs pointing to chance variables are consistent with the partial order determined by information constraints, then again no divisions are necessary. The American put option is an example of this genre. Also, Markov reward processes, where the arcs always point forward in time (from state $S(t)$ to state $S(t+1)$), are a class of problems where divisions are not required.

6.1. Limitations

For more general decision problems where divisions are needed to solve a problem using local computation, the method described in this paper will not work. The family of MOP functions is not closed under the division operation. Thus, if we divide a MOP function by another MOP function, the resulting function may not be a MOP, in which case there are no guarantees that we can integrate such functions in closed form. The *Pigs* problem, discussed by Lauritzen and Nilsson (2001), is an example of a problem of this type (requiring divisions for solution using local computation). In general, if we have an additive factorization of the joint utility function, and arc reversals are necessary for solution, then such problems cannot be solved using local computation by using MOPs.

Our method based on MOPs inherits all the problems and issues that are inherent with the MOP method. First, we need to find MOP approximations of PDFs and utility functions. We can find MOP approximations by using Lagrange interpolating polynomials with Chebyshev points, but it needs manual interventions regarding the location of knots that make up the pieces. Currently, we have some heuristics (mode, inflection points, equal width, etc.), but no theory for this decision. Except for this issue, we can automate the process of finding MOP approximations of conditional PDFs, and the process of solving hybrid IDs containing deterministic conditionals. Lagrange interpolating polynomials for one- and higher-dimensional functions can be easily found using commercial software (such as Mathematica, Maple, Matlab, etc.). The LIP method does not require that the function being approximated be differentiable. The theory of Chebyshev points exists for one- and two-dimensional functions. Also, we can use a MOP approximation of the one-dimensional standard normal PDF to construct MOP approximations of higher-dimensional CLG PDFs. However, constructing tractable MOP approximations of high-dimensional non-CLG PDFs (such as a three-dimensional log-normal PDF) can be a challenge.

6.2. Future Work

How does the MOP method compare with the discretization and Monte Carlo methods? This is an important question that needs to be answered and

for which we do not have answers. At this stage, we note that discretization has only been studied for one-dimensional PDFs. Although this can be naïvely applied to multidimensional conditional PDFs, the quality of the resulting approximation has not been studied. For the conditional PDFs in the American put option problem, one could, e.g., find a three-point discrete approximation of the PDF of S_1 ; a three-point discrete approximation of the conditional PDF of $S_2 | s_1$ for each of the three values of S_1 , resulting in up to $3^2 = 9$ distinct values of S_2 ; a three-point discrete approximation of the PDF of $S_3 | s_2$ for each of the nine distinct values of S_2 , resulting in up to $3^3 = 27$ distinct values of S_3 ; etc. Clearly, such a strategy is not tractable for many stages. Also, we cannot imagine obtaining a strategy as detailed as the one we obtain in stage 1 (exercise the option in stage 1 if the observed value of S_1 is less than 29.16, and hold otherwise) from a discretized model with only three possible values of S_1 . Finally, we note that Markov chain Monte Carlo methods wouldn't converge for a probability model that includes deterministic conditionals.

How close is the approximate solution found by using our method to the true answer? This is another important question for which we do not have answers. We note that the errors in the MOP approximation of conditional PDFs can be quantified using measures such as the Kullback and Liebler (1951) divergence and maximum absolute deviation between the MOP approximation and the target PDF (Shenoy 2012). In terms of these measures, the approximations have very small errors. For example, the Kullback and Liebler (1951) divergence between the standard normal PDF truncated to $(-3, 3)$ and the two-piece, three-degree MOP approximation described in Equation (8) is 0.009, and the maximum absolute deviation between the two functions is 0.014. However, we do not know how these errors influence the errors in the optimal strategy and the errors in the maximum expected utility. This is a topic that needs further research.

What is the size of decision problems that can be solved by our method? This is yet another important question for which we do not have answers. We plan to solve the American put option problem by gradually increasing the number of stages and observing where the method breaks down, if at all. The

main problem here is computing a MOP approximation of the conditional for $S_j | s_{j-1}$. As we change the number of stages, we need to recompute all of the MOP approximations of the conditional PDFs. This is another topic for further research.

Acknowledgments

The authors are grateful for comments from three anonymous reviewers from the Uncertainty in Artificial Intelligence 2010 conference and comments from two anonymous reviewers and an associate editor of *Decision Analysis*. A short version of this paper appeared in the proceedings of the Uncertainty in Artificial Intelligence 2010 conference (Li and Shenoy 2010).

Appendix. Properties of Dirac Delta Functions

Two basic properties of Dirac delta functions are as follows (see, e.g., Dirac 1927, 1958; Hoskins 1979; Kanwal 1998).

1. (*Sampling*) If $f(x)$ is any function that is continuous in the neighborhood of a , then

$$\int_{-\infty}^{\infty} f(x) \delta(x - a) dx = f(a). \quad (29)$$

2. (*Rescaling*) If $g(x)$ has real (noncomplex) zeros at a_1, \dots, a_n and is differentiable at these points, and $g'(a_i) \neq 0$ for $i = 1, \dots, n$, then

$$\delta(g(x)) = \sum_{i=1}^n \frac{\delta(x - a_i)}{|g'(a_i)|}.$$

A more extensive list of properties of the Dirac delta function that are relevant for uncertain reasoning can be found in Cinicioglu and Shenoy (2009).

References

- Bielza C, Müller P, Ríos Insua D (1999) Decision analysis by augmented probability simulation. *Management Sci.* 45(7):995–1007.
- Black F, Scholes M (1973) The pricing of options and corporate liabilities. *J. Political Econom.* 81(3):1323–1352.
- Burden RL, Faires JD (2010) *Numerical Analysis*, 9th ed. (Brooks Cole, Boston).
- Cano A, Gómez M, Moral S (2006) A forward-backward Monte Carlo method for solving influence diagrams. *Internat. J. Approximate Reasoning* 42(1–2):119–135.
- Charnes JM, Shenoy PP (2004) Multi-stage Monte Carlo method for solving influence diagrams using local computation. *Management Sci.* 50(3):405–418.
- Cinicioglu EN, Shenoy PP (2009) Arc reversals in hybrid Bayesian networks with deterministic variables. *Internat. J. Approximate Reasoning* 50(5):763–777.
- Cobb BR (2007) Influence diagrams with continuous decision variables and non-Gaussian uncertainties. *Decision Anal.* 4(3):136–155.
- Cobb BR, Shenoy PP (2005a) Hybrid Bayesian networks with linear deterministic variables. Bacchus F, Jaakkola T, eds. *Uncertainty in Artificial Intelligence: Proc. 21st Conf.* (AUAI Press, Corvallis, OR), 136–144.
- Cobb BR, Shenoy PP (2005b) Nonlinear deterministic relationships in Bayesian networks. Godo L, ed. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 8th European Conf., ECSQARU 2005*. Lecture Notes in Artificial Intelligence 3571 (Springer, Berlin), 27–38.
- Cobb BR, Shenoy PP (2008) Decision making with hybrid influence diagrams using mixtures of truncated exponentials. *Eur. J. Oper. Res.* 186(1):261–275.
- Cobb BR, Shenoy PP, Rumi R (2006) Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. *Statist. Comput.* 16(3):293–308.
- Dirac PAM (1927) The physical interpretation of the quantum dynamics. *Proc. Royal Soc. London, Ser. A* 113(765):621–641.
- Dirac PAM (1958) *The Principles of Quantum Mechanics*, 4th ed. (Oxford University Press, London).
- Geske R, Johnson HE (1984) The Americal put option valued analytically. *J. Finance* 39(5):1511–1524.
- Hertz DB (1964) Risk analysis in capital investment. *Harvard Bus. Rev.* 42(1):95–106.
- Hoskins RF (1979) *Generalised Functions* (Ellis Horwood, Chichester, UK).
- Howard RA (1971) Proximal decision analysis. *Management Sci.* 17(9):845–879.
- Howard RA, Matheson JE (1984) Influence diagrams. Howard RA, Matheson JE, eds. *Readings on the Principles and Applications of Decision Analysis*, Vol. II. (Strategic Decisions Group, Menlo Park, CA), 719–762.
- Howard RA, Matheson JE (2005) Influence diagrams. *Decision Anal.* 2(3):127–143.
- Kanwal RP (1998) *Generalized Functions: Theory and Technique*, 2nd ed. (Birkhäuser, Boston).
- Keefer DL, Bodily SE (1983) Three-point approximations for continuous random variables. *Management Sci.* 29(5):595–609.
- Keelin TW, Powley BW (2011) Quantile-parameterized distributions. *Decision Anal.* 8(3):206–219.
- Kozlov AV, Koller D (1997) Nonuniform dynamic discretization in hybrid networks. Geiger D, Shenoy PP, eds. *Uncertainty in Artificial Intelligence: Proc. 13th Conf.* (Morgan Kaufmann, San Francisco), 302–313.
- Kschischang FR, Frey BJ, Loeliger H-A (2001) Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory* 47(2):498–519.
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann. Math. Statist.* 22:76–86.
- Lauritzen SL, Nilsson D (2001) Representing and solving decision problems with limited information. *Management Sci.* 47(9):1235–1251.
- Li Y, Shenoy PP (2010) Solving hybrid influence diagrams with deterministic variables. Grünwald P, Spirtes P, eds. *Uncertainty in Artificial Intelligence: Proc. 26th Conf.* (AUAI Press, Corvallis, OR), 322–331.
- Miller AC III, Rice TR (1983) Discrete approximations of probability distributions. *Management Sci.* 29(3):352–362.
- Moral S, Rumi R, Salmerón A (2001) Mixtures of truncated exponentials in hybrid Bayesian networks. Benferhat S, Besnard P, eds. *Symbolic and Quant. Approaches to Reasoning with Uncertainty: 6th Eur. Conf., ECSQARU-2001*. Lecture Notes in Artificial Intelligence 2143 (Springer, Berlin), 156–167.
- Olmsted SM (1983) On representing and solving decision problems. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University, Stanford, CA.
- Ortiz LE, Kaelbling LP (2000) Sampling methods for action selection in influence diagrams. Kautz H, Porter B, eds. *Proc. 17th National Conf. Artificial Intelligence* (AAAI Press, Menlo Park, CA), 378–385.

- Poland WB (1994) Decision analysis with continuous and discrete variables: A mixture distribution approach. Ph.D. thesis, Department of Engineering-Economic Systems, Stanford University, Stanford, CA.
- Poland WB, Shachter RD (1993) Mixtures of Gaussians and minimum relative entropy techniques for modeling continuous uncertainties. Heckerman D, Mamdani A, eds. *Uncertainty in Artificial Intelligence: Proc. 9th Conf.* (Morgan Kaufmann, San Francisco), 183–190.
- Shachter RD, Kenley CR (1989) Gaussian influence diagrams. *Management Sci.* 35(5):527–550.
- Shenoy PP (1992) Valuation-based systems for Bayesian decision analysis. *Oper. Res.* 40(3):463–484.
- Shenoy PP (2012) Two issues in using mixtures of polynomials for inference in hybrid Bayesian networks. *Internat. J. Approximate Reasoning*, ePub ahead of print February 7, <http://dx.doi.org/10.1016/j.ijar.2012.01.008>.
- Shenoy PP, Shafer G (1990) Axioms for probability and belief-function propagation. Shachter RD, Levitt T, Lemmer JE, Kanal LN, eds. *Uncertainty in Artificial Intelligence 4* (North-Holland, Amsterdam), 169–198.
- Shenoy PP, West JC (2011a) Extended Shenoy-Shafer architecture for inference in hybrid Bayesian networks with deterministic conditionals. *Internat. J. Approximate Reasoning* 52(6):805–818.
- Shenoy PP, West JC (2011b) Inference in hybrid Bayesian networks using mixtures of polynomials. *Internat. J. Approximate Reasoning* 52(5):641–657.
- Shenoy PP, Rumí R, Salmerón A (2011) Some practical issues in inference in hybrid Bayesian networks with deterministic conditionals. *IEEE Proc. 2011 Eleventh Internat. Conf. Intelligent Systems Design Appl. (ISDA-11)* (IEEE Research Publishing Services, Piscataway, NJ), 605–610.
- Smith JE (1993) Moment methods for decision analysis. *Management Sci.* 39(3):340–358.
- Tatman JA, Shachter RD (1990) Dynamic programming and influence diagrams. *IEEE Trans. Systems, Man and Cybernetics* 20(2):365–379.
- Xu Y (1996) Lagrange interpolation on Chebyshev points of two variables. *J. Approximation Theory* 87(2):220–238.