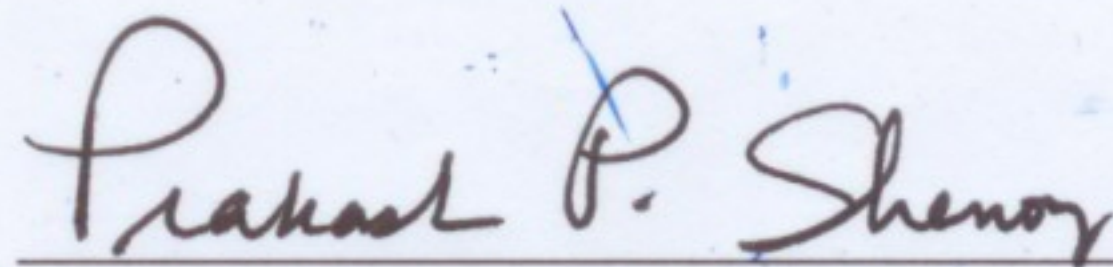ON SOLVING STOCHASTIC PERT NETWORKS
AND
USING RFIDs FOR OPERATIONS MANAGEMENT

BY
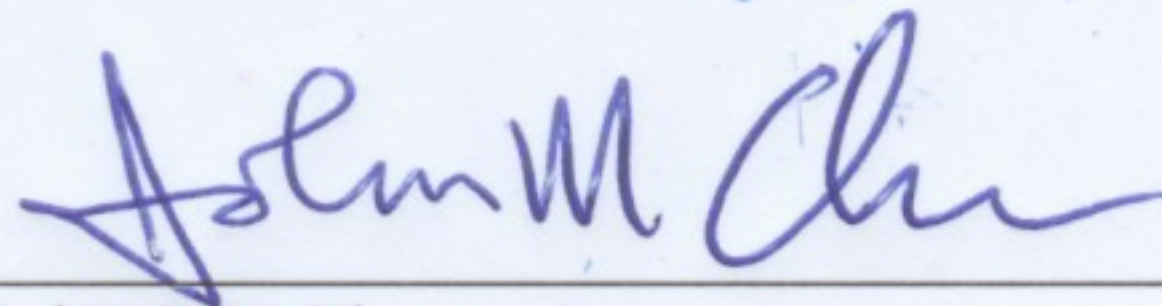
Esma Nur Cinicioglu
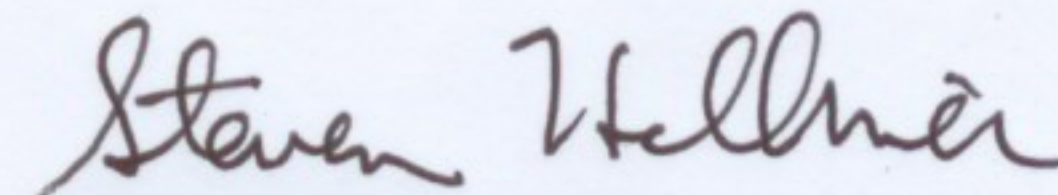B.S., Marmara University, 2002

Submitted to the School of Business and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
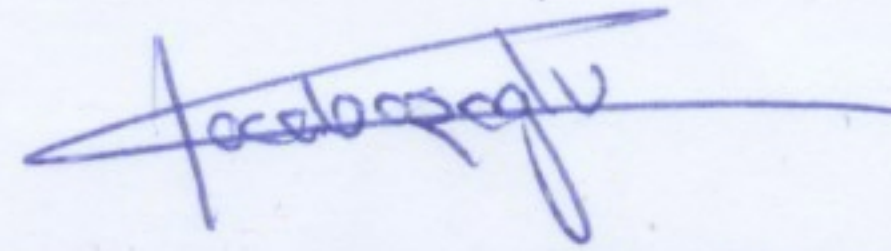Doctor of Philosophy.

Prakash Shenoy (Chair)
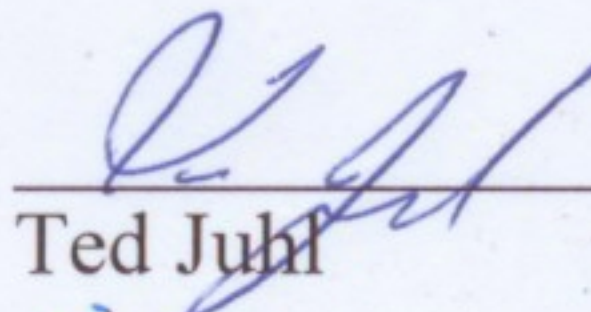
John M. Charnes

Steve Hillmer

Canan Kocabasoglu-Hillmer

Ted Juhl

Date Defended: April 25, 2008

UMI Number: 3297865

Copyright 2008 by
Cinicioglu, Esma Nur

All rights reserved.

UMI®
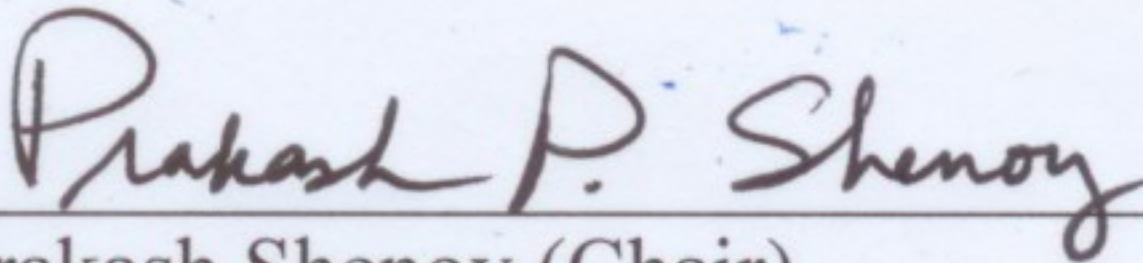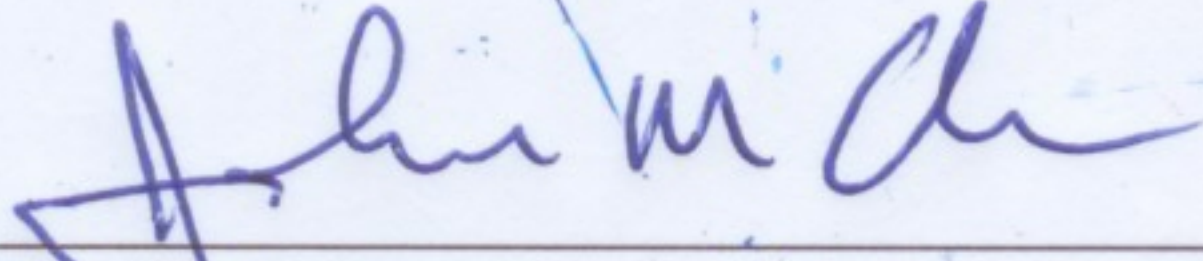
The Dissertation Committee for Esma Nur Cinicioglu certifies
that this is the approved version of the following dissertation:

ON SOLVING STOCHASTIC PERT NETWORKS
AND
USING RFIDs FOR OPERATIONS MANAGEMENT

Committee:

_____
Prakash Shenoy (Chair)

_____
John M. Charnes

_____
Steve Hillmer

_____
Canan Kocabasoglu-Hillmer

_____
Ted Juhl

Date approved: 4/25/2008

i

Abstract

The current methods used to solve stochastic PERT networks overlook the true distribution of the maximum of two distributions and thus fail to compute an accurate estimation of the project completion time. This dissertation presents two different methods to solve stochastic PERT networks. With each method, both by using mixtures of Gaussians and also by using mixtures of truncated exponentials, the distribution of the maximum of two distributions can be approximated accurately.

In the first method a PERT network is first transformed into a MoG Bayesian network and then Lauritzen-Jensen algorithm is used to make inferences in the resulting MoG Bayesian network. The transformation process involves approximating non-Gaussian distributions using MoG's, finding maximum of two distributions 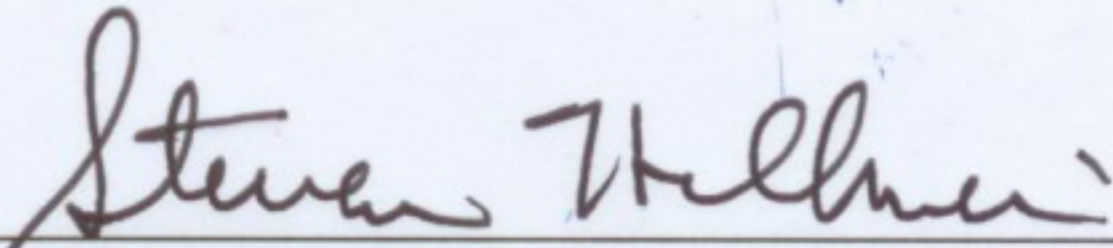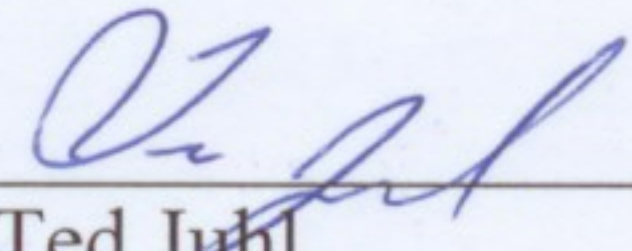using MoG's. As PERT networks are transformed into MoG Bayesian networks arc reversals may also become necessary since MoG Bayesian networks does not allow discrete variables to have continuous parents. This dissertation presents arc reversals in hybrid Bayesian networks with deterministic variables between every possible pair of variables.

In the second stage of the research MTE Bayesian networks are introduced as an alternative for solving stochastic PERT networks. We demonstrated the easy applicability of MTE potentials by finding the marginal probability distribution of a PERT example using MTE's. This calculation process involves the conversion of the PERT network into a PERT Bayes net, transformation of the PERT Bayes net into a MTE network and finally propagation of the MTE potentials using the Shenoy-Shafer

architecture. Finding the distribution of the maximum of two distributions using MTE's is described as an operation necessary to propagate in MTE PERT networks.

The second essay of this dissertation discusses a potential application of radio frequency identification (RFID) and collaborative filtering for targeted advertising in grocery stores. Every day hundreds of items in grocery stores are marked down for promotional purposes. Whether these promotions are effective or not depends primarily on whether the customers are aware of them or not and secondarily whether the products on promotion are products in which the customer will be interested. Currently, the companies are relatively incapable of influencing the customers' decision-making process while they are shopping. However, the capabilities of RFID technology enable us to transfer the recommendation systems of e-commerce to grocery stores. In our model, using RFID technology, we get real time information about the products placed in the cart during the shopping process. Based on that information we inform the customer about those promotions in which the customer is likely to be interested in. The selection of the product advertised is a dynamic decision making process since it is based on the information of the products placed inside the cart while customer is shopping. Collaborative filtering is used for the identification of the advertised product and Bayesian networks will be used for the application of collaborative filtering. We are assuming a scenario where all products have RFID tags, and grocery carts are equipped with RFID readers and screens that would display the relevant promotions. We present our model first using the data set available for the Netflix prize competition. As the second stage of the research we use

grocery data set and develop a new heuristic to select the products to be used in the

Bayesian network created.

Acknowledgements

First of all, I would like to express my sincere appreciation to Prakash Shenoy, my advisor. Having him as my mentor I always considered myself as blessed. Throughout these years I worked with him, I gained invaluable knowledge as a researcher and I am really grateful for that. But beyond that my real gain I believe, was the opportunity to observe him as my role model in life who sets an example by showing brilliancy both in terms of research and also in terms of excellent personality.

I would like to extend my special thanks to Steve Hillmer and to John Charnes who supported me through my whole PhD study. Their guidance, judgment and expertise have been invaluable. I would like to express my appreciation to Canan Kocabasoglu-Hillmer for her support both academically and personally. I am grateful to Ted Juhl for serving on my committee, for providing insightful comments and suggestions, and for his inspirational teaching.

Foremost, I would like to thank my whole family who supported me by any means and encouraged me any time of any day during my whole PhD life. With all their love, encouragement and mentorship, I believe in addition to their existing PhDs in their fields, they also deserve to have a PhD for being the best family possible.

TABLE OF CONTENTS

1 REVIEW OF PROJECT EVALUATION AND REVIEW TECHNIQUE

1.1 PERT Networks

Program Evaluation and Review Technique (PERT) was invented in 1958 for the POLARIS missile program by the Program Evaluation branch of the Special Projects Office of the U. S. Navy, assisted by consultants from Booz, Allen and Hamilton [Malcolm *et al*. 1959]. A parallel technique called *Critical Path Method* (CPM) was invented around the same time by Kelley and Walker [1959]. Both PERT and CPM are project management techniques whose main goal is to manage the completion time of a large project consisting of many activities with precedence constraints, i.e., constraints that specify other activities that need to be completed prior to starting an activity.

In PERT, a project is represented by a directed acyclic network where the nodes represent duration of activities and the arcs represent precedence constraints. In classical PERT, duration of activities are assumed to be known constants, and the task is to identify a "critical path" from start-time to finish-time such that the project completion time is the sum of the duration of the activities on the critical path. These activities are called *critical*, since a project could be delayed if these activities were not completed in the scheduled time. In stochastic PERT, activities are considered as random variables with probability distributions, and the main task is to compute the marginal probability distribution of the project completion time.

The problem of computing the marginal probability distribution of the project completion time is a difficult problem. Thus many approximate techniques have been developed. A classic solution proposed by Malcolm *et al*. [1959] is to assume that all activities are independent random variables and that each activity has an approximate beta distribution parameterized by three parameters: mean time m, minimum (optimistic) completion time a, and maximum (pessimistic) completion time b. The expected duration of each activity is then approximated by $(a + 4m + b)/6$, and its variance is approximated by $(b - a)^2/36$. Using the expected duration times, the critical path is computed using the classical deterministic method. Assuming independence, the mean and variance of the distribution of the project completion time is then approximated as the sum of the expected durations and the sum of variances of the activities on a critical path.

Another approximation is to assume that all activity durations are independent and have the Gaussian distribution [Sculli 1983]. The completion time of an activity i is given by $C_i = \text{Max}\{C_j \mid j \in \Pi(i)\} + D_i$, where C_j denotes the completion time of activity j, D_j denotes the duration of activity j, and $\Pi(i)$ denotes the parents (immediate predecessors) of activity i. The maximum of two independent Gaussian random variables is not Gaussian. However, the distribution of C_i is assumed to be Gaussian with the parameters estimated from the parameters of the parent activities. Depending on the values of the parameters, this assumption can lead to large errors.

Kulkarni and Adlakha [1986] compute the distribution and moments of project completion time assuming that the activity durations are independent and having the

exponential distribution with finite means. They call such stochastic PERT networks Markov networks.

If we don't assume independence of activity durations, the problem of computing the marginal distribution of the project completion time becomes computationally intractable for large projects. One solution to this problem is to use Monte Carlo techniques with variance reduction techniques to estimate the distribution of project completion time or its moments [Van Slyke 1963, Burt and Garman 1971, Garman 1972, Sigal *et al*. 1979, Fishman 1985]. Another solution is to provide lower bounds for the expected project completion time [see e.g., Elmaghraby 1967, Fulkerson 1962, and Robillard 1976]. Elmaghraby [1977] provides a review of Monte Carlo and bounding techniques.

Jenzarli [1995] suggests the use of Bayesian networks to model the dependence between activity durations and completions in a project. Following Jenzarli, we will first transform PERT networks into PERT Bayesian networks. Afterward we will approximate it by a mixture of Gaussians (MoG) Bayes net, and then use the Lauritzen-Jensen algorithm to make exact inferences in the MoG Bayes net.

The next sections provide information about Bayesian networks and how to transform a stochastic PERT network into a Bayesian network.

1.2   Bayesian Networks

A Bayesian network is a directed acyclic graph where nodes represent the variables and the arcs represent the conditional independencies between the variables. If there

is a directed arc from a variable X_1 to a variable X_2 then we call X_1 as the parent of X_2 and X_2 as the child of X_1. Each variable in a Bayesian network $X_1, \ldots, X_N$ possess a probability distribution given its parents and the product of these conditional probability distributions constitute the joint probability distribution of the network.

$$P(X_1, \ldots, X_N) = \prod_{i=1}^{N} P(X_i \mid Pa(X_i)) \qquad (1.1)$$

Jenzarli [1995] suggests the use of Bayesian networks to model the dependence between activity durations and completions in a project. However, such Bayesian networks are difficult to solve exactly since they may contain a mix of discrete and continuous random variables. The solution recommended by Jenzarli is to use Markov chain Monte Carlo techniques to estimate the marginal distribution of project completion time.

Bayesian networks containing a mix of discrete (with a countable number of outcomes) and continuous (real-valued) chance variables are called Hybrid Bayesian networks. Shenoy [2006] describes a new technique for "exact" inference in hybrid Bayesian networks using mixture of Gaussians. This technique consists of transforming a general hybrid Bayesian network to a mixture of Gaussians Bayesian network.

In chapter 2, we explore the use of exact inference in hybrid Bayesian networks using mixture of Gaussians proposed by Shenoy [2006] to compute the exact marginal distribution of project completion time. Activities durations can have any distribution, and may not be all independent. We model dependence between

activities using a Bayesian network as suggested by Jenzarli [1995]. We approximate non-Gaussian conditional distributions by mixture of Gaussians, and we reduce the resulting hybrid Bayesian network to a mixture of Gaussian Bayesian networks. Such hybrid Bayesian networks can be solved exactly using the algorithm proposed by Lauritzen and Jensen [2001], which is implemented in Hugin, a commercially-available software package. In the following section we illustrate our approach using a small PERT network with five activities.

1.3   Representing Stochastic PERT Network as Bayesian Network

Consider a PERT network as shown in Figure 1.1 with five activities, A_1, …, A_5. S denotes project start time, and F denotes project completion time. The directed arrows in a PERT network denote precedence constraints. The precedence constraints are as follows. A_3 and A_5 can only be started after A_1 is completed, and A_4 can only be started after A_2 and A_3 are completed. The project is completed after all five activities are completed.
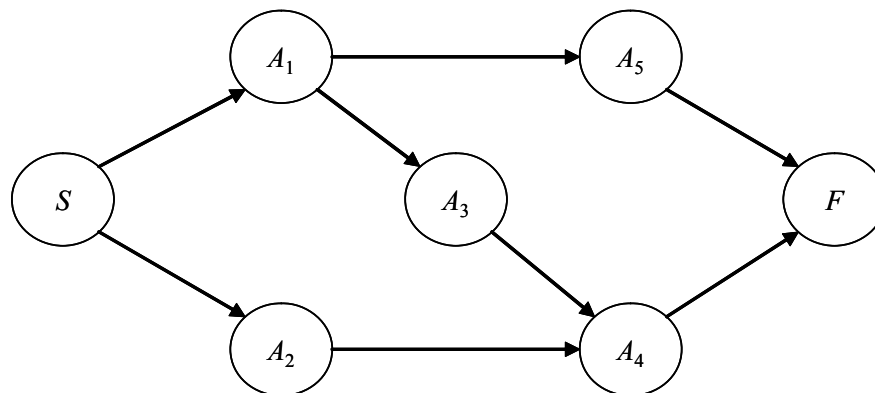
Figure 1.1: A stochastic PERT network with five activities.

Using the technique described in Jenzarli [1995], we will describe the dependencies of the activities by a Bayesian network. Let D_i denote the duration of activity i, and let C_i denote the earliest completion time of activity i. Let C_{23} denote earliest completion time of activities 2 and 3. Since our goal is to compute the marginal distribution of the earliest completion time of the project, we will assume that each activity will be started as soon as possible (after completion of all preceding activities). Also, we assume that $S = 0$ (with probability 1).

The interpretation of PERT networks as Bayes Nets allows us to depict the activity durations that are dependent on each other. For instance, in the current example durations of activities 1 and 3 and durations of activities 2 and 4 are positively correlated. Considering the dependencies between the activities, we convert the PERT network to a Bayes net following three basic steps. First activity durations are replaced with activity completion times, second activity durations are added with an arrow from D_i to C_i so that each activity is represented by two nodes. However, notice that the activities 1 and 2 are represented just by their durations, as D_1 and D_2. The reason for that is that they are starting activities and since they do not have any predecessors, the completion times of the activities will be the same as their durations. As the last step we represent the dependence between durations by arc, so an arc is added from D_1 to D_3 and from D_2 to D_4. The resulting PERT Bayes net representation of the PERT network is illustrated in Figure 1.2 below.
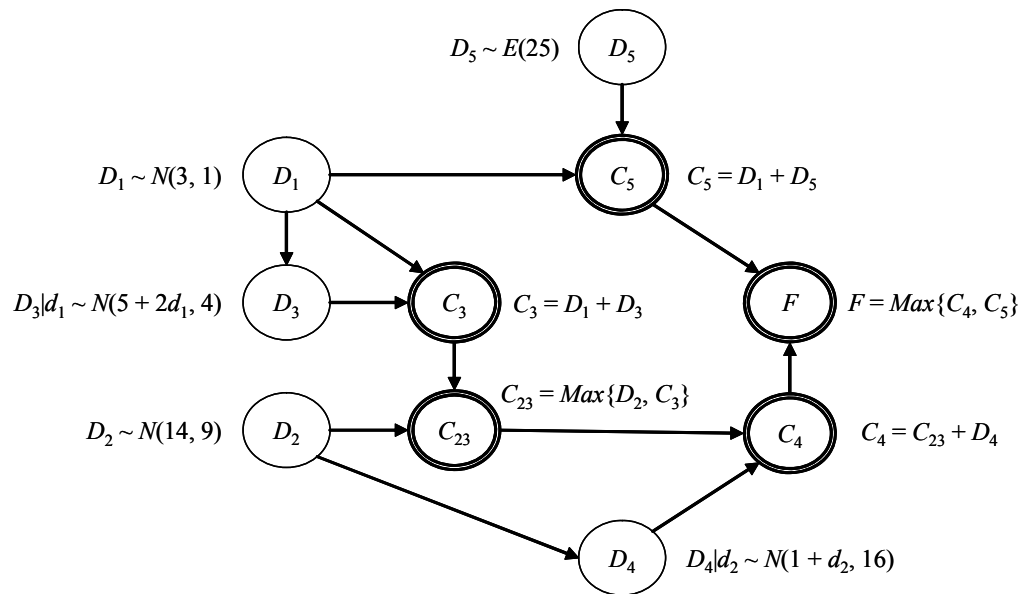
Figure 1.2: A Bayes net representation of the dependencies of the activities in the

PERT network of Figure 1.1.

1.4   Summary and Conclusions

This chapter illustrates methods that are used and proposed in the literature for solving stochastic PERT networks. In this scope, we talked about the problems involved with these methods and discussed the restrictive assumptions made in the literature.

We demonstrated how to represent a PERT network as a Bayesian network using Jenzarli's method which allows us to model the dependencies between the activities and serves as a base for the next step of this research where we explore the use of exact inference in hybrid Bayesian networks using mixtures of Gaussians proposed by Shenoy [2006] to compute the exact marginal distribution of project completion time.

2 SOLVING STOCHASTIC PERT NETWORKS USING MIXTURES OF GAUSSIANS

2.1 Mixtures of Gaussians Bayesian Networks

Mixtures of Gaussians (MoG) hybrid Bayesian networks were initially studied by Lauritzen [1992]. These are Bayesian networks with a mix of discrete and continuous variables. The discrete variables cannot have continuous parents, and all continuous variables have the so-called conditional linear Gaussian distributions. This means that the conditional distributions at each continuous node have to be Gaussian such that the mean is a linear function of its continuous parents, and the variance is a constant. MoG Bayesian networks have the property that for each instantiation of the discrete variables, the joint conditional distribution of the continuous variables is multivariate Gaussian. Hence the name 'mixtures of Gaussians.' An example of a MoG Bayesian network is as shown in Figure 2.1.
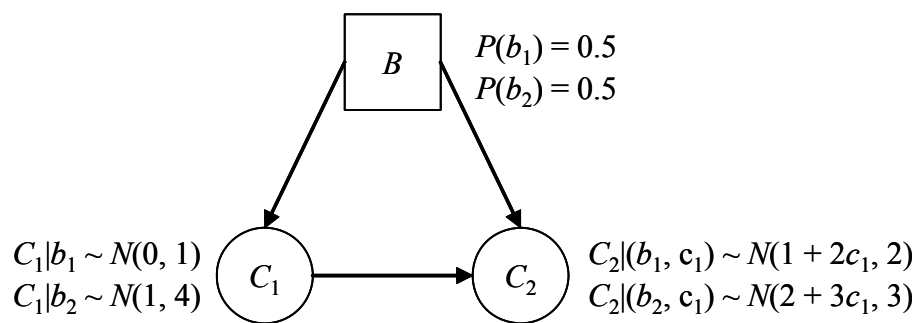
Figure 2.1: An example of a MoG Bayes net.

Consider the Bayes Net shown in Figure 1.2. It is not a MoG Bayesian network since D_5 has a non-Gaussian distribution, and C_{23} and F have a non-linear conditional Gaussian distribution.

Using the method described in Shenoy [2006] the non-Gaussian distributions and the non-linear Gaussian distributions can be approximated using Mixtures of Gaussians. In the process of doing so, we may create discrete variables with continuous parents. In this case, arc reversals become necessary to convert the resulting hybrid Bayesian network to a MoG Bayesian network. In the next section arc reversals between every possible kind of pairs of variables will be described. Following that it will be explained how we can approximate a non-Gaussian distribution by a MoG distribution, and how we can approximate a max deterministic function by a MoG distribution, for which the use of arc reversals will be necessary.

2.2   Arc Reversals in Hybrid Bayesian Networks

If we have a general hybrid Bayesian network containing a discrete variable with continuous parents, then one method of transforming such a network to a MoG Bayesian network is to do arc reversals. If a continuous variable has a non-CLG distribution, then we can approximate it with a MoG distribution. In the process of doing so, we may create a discrete variable with continuous parents. In this case, arc reversals are again necessary to convert the resulting hybrid Bayesian network to a MoG Bayesian network. Arc reversals are also used to solve influence diagrams, which are Bayesian networks with decision and utility nodes. Although there are no

exact algorithms to solve hybrid influence diagrams (containing a mix of discrete and continuous chance variables), a theory of arc reversals is potentially useful in this endeavor.

Arc reversals were pioneered by Olmsted [1984] for solving discrete influence diagrams. They were further studied by Shachter [1986, 1988, 1990] for solving discrete influence diagrams, finding posterior marginals in Bayesian networks, and for finding relevant sets of variables for an inference problem. Kenley [1986] generalized arc reversals in influence diagrams with continuous variables having conditional linear Gaussian distributions (see also Shachter and Kenley [1989]). Poland (1994) further generalized arc reversals in influence diagrams with Gaussian mixture distributions. Although there are no exact algorithms to solve general hybrid influence diagrams (containing a mix of discrete and continuous chance variables), a theory of arc reversals is potentially useful in this endeavor.

Hybrid Bayesian networks containing deterministic variables with continuous parents pose a special problem since the joint density for all continuous variables does not exist. Thus, a method for propagating density potentials would need to be modified to account for the non-existence of the joint density.[Cobb and Shenoy 2005, 2006b, 2007]

2.2.1   Notation

In this section we will describe the notation and definitions used in this research. The notation and definitions are adapted from Cobb and Shenoy [2005].

Variables and States. We are concerned with a finite set V of variables. Each variable $X \in V$ is associated with a set of its possible *states* denoted by Ω_X. If Ω_X is a countable set, finite or infinite, we say X is *discrete*, and depict it by a rectangular node in a graph; otherwise X is said to be *continuous* and is depicted by an oval node.

In a Bayesian network, each variable is associated with a conditional distribution for each state of its parents. A conditional distribution function associated with a continuous variable is said to be *deterministic* if its values are in units of probability mass. For simplicity, we will refer to continuous variables with non-deterministic conditionals as continuous, and continuous variables with deterministic conditionals as deterministic. Deterministic variables are represented as oval nodes with a double border in a graph.

We will assume that the state space of continuous variables is the set of real numbers (or some subset of it) and that the states of a discrete variable are symbols. If $r \subseteq V$, then $\Omega_r = \times \{\Omega_X \mid X \in V\}$.

Potentials. In a Bayesian network, each variable is associated with a conditional probability function given its parents and these are represented by functions called *potentials*. If X is discrete, it is associated with a *discrete* potential. Formally, suppose r is a set of variables that contains a discrete variable. A discrete potential ρ for r is a function $\rho: \Omega_r \to [0, 1]$. The values of discrete potentials are in units of probability mass.

Although the domain of the function ρ is Ω_r, for simplicity, we will refer to r as the *domain* of ρ. Thus the domain of a potential representing the conditional

probability mass function associated with some variable X in a Bayesian network is always the set $\{X\} \cup pa(X)$, where $pa(X)$ denotes the set of parents of X. The set $pa(X)$ may contain continuous variables.

Continuous non-deterministic variables are typically associated with conditional density functions, which are represented by functions called *density potentials*. Formally, suppose r is a set of variables that contains a continuous variable. A density potential ρ for r is a function $\rho: \Omega_r \rightarrow P^+$. The values of density potentials are in units of probability density.

Deterministic variables are associated with conditional distributions containing equations and whose values are in units of probability mass. We will call such functions *equation* potentials. Formally, suppose $x = r \cup s$ is a set of variables containing some discrete variables r and some continuous variables s. We assume $s \neq \varnothing$. An equation potential ξ for x is a function $\xi: \Omega_x \rightarrow [0, 1]$ such that $\xi(r, s)$ is of the form $\Sigma\{p_i(r, s) [Z = g_{r,i}(s\backslash\{Z\})](s) \mid i = 1, \ldots, n\}$, where $[Z = g_{r,i}(s\backslash\{Z\})]$ are indicator functions such that $[Z = g_{r,i}(s\backslash\{Z\})](s) = 1$ if $z = g_{r,i}(s\backslash\{z\})$, and $= 0$ otherwise, and $p_i(r, s)$ have units of probability mass, for all $i = 1, \ldots, n$. The values of equation potentials are in units of probability mass. Suppose Y is a deterministic variable with continuous parent X, and suppose that the deterministic relationship is $Y|x = x$ with probability ½ and $Y|x = 0$ with probability ½. This distribution is represented by the equation potential $½ [Y = X] + ½ [Y = 0]$ for $\{X, Y\}$, where $[Y = X](x, y) = 1$ if $y = x$, and $= 0$ otherwise, and $[Y = 0](y) = 1$ if $y = 0$, and $= 0$ otherwise. Notice that our definition of deterministic variables is slightly more

inclusive than the usual definition of variables whose conditional distributions have zero variances. It includes conditional distributions as above with probability masses, which have non-zero variances.

Both density and equation potentials are special instances of a broader class of potentials called continuous potentials. Suppose z is a set of variables containing a continuous variable. Then a *continuous potential* ζ for z is a function $\zeta: \Omega_x \rightarrow [0, 1] \cup P^+$. The values of ζ may have units of probability mass (in $[0, 1]$) or probability density (in P^+). For example, consider a continuous variable X with a mixed distribution: a probability mass of 0.5 at $X = 0$, and a probability density of $0.5\,f$, where f is a probability density function whose values are in units of probability density. This mixed distribution can be represented by a continuous potential ξ for $\{X\}$ as follows: $\xi(x) = 0.5\,[X = 0](x) + 0.5\,f(x)$. The first part has units of probability mass and the second part has units of probability density. When we wish to be explicit about this, we will write $\xi(x) = 0.5\,[X = 0](x)\,(m) + 0.5\,f(x)\,(d)$.

As we will see shortly, the combination of two density potentials is a density potential, the combination of two equation potentials is an equation potential, and the combination of two continuous potential is a continuous potential. Also, continuous potentials can result from the marginalization and division operations. These operations will be defined shortly.

$$X|(d_1, \ z) = z \text{ with prob. } 1$$
$$X|(d_2, \ z) = 0 \text{ with prob. } 1$$
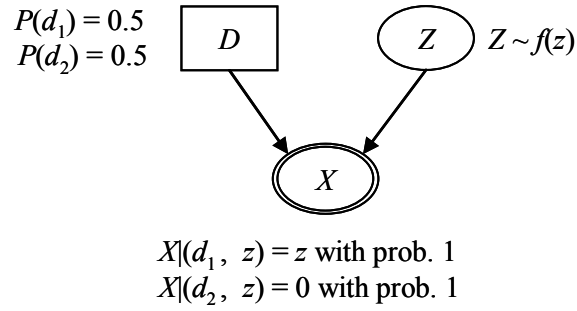
Figure 2.2: A Bayesian network with a discrete, a continuous and a deterministic

variable

Consider the BN given in Figure 2.2. In this BN, D is a discrete variable with two

states d_1 and d_2 with the discrete distribution $P(d_1) = 0.5$, $P(d_2) = 0.5$. Z is a

continuous variable with a probability density function (PDF) f. X is a deterministic

variable with the conditional distribution $X|(d_1, z) = z$ with probability 1, and $X|(d_2, z)$

$= 0$ with probability 1. Let δ denote the discrete potential for $\{D\}$ associated with D.

Then, $\delta(d_1) = 0.5$ and $\delta(d_2) = 0.5$, Let ζ be the density potential for $\{Z\}$ associated

with Z. Then, $\zeta(z) = f(z)$. Let ξ denote the equation potential for $\{D, Z, X\}$ associated

with X. Then, $\xi(d_1, z, x) = [X = Z](z, x)$ and $\xi(d_2, z, x) = [X = 0](x)$.

Next, we define three operations associated with potentials, combination,

marginalization, and division. Before we define combination of potentials, we need to

define projection of states. Suppose y is a state of variables in r, and suppose $s \subseteq r$.

Then the *projection* of y to s, denoted by $y^{\downarrow s}$ is the state of s obtained from y by

dropping states of $r \backslash s$. Thus, $(w, x, y, z)^{\downarrow \{W, X\}} = (w, x)$, where $w \in \Omega_W$, and $x \in \Omega_X$. If s

$= r$, then $y^{\downarrow s} = y$.

Combination. Suppose α is a potential (discrete or continuous) for a, and β is a potential (discrete or continuous) for b. Then the *combination* of α and β, denoted by $\alpha \otimes \beta$, is the potential for $a \cup b$ obtained from α and β by pointwise multiplication, i.e., $(\alpha \otimes \beta)(\boldsymbol{x}) = \alpha(\boldsymbol{x}^{\downarrow a})\, \beta(\boldsymbol{x}^{\downarrow b})$ for all $\boldsymbol{x} \in \Omega_{a \cup b}$. If α and β are both discrete potentials, then $\alpha \otimes \beta$ is a discrete potential. If α and β are both density potentials, then $\alpha \otimes \beta$ is a density potential. If α and β are both equation potentials, then $\alpha \otimes \beta$ is an equation potential. And if α and β are both continuous potentials, then $\alpha \otimes \beta$ is a continuous potential.

Combination of potentials (discrete or continuous) is commutative ($\alpha \otimes \beta = \beta \otimes \alpha$) and associative (($\alpha \otimes \beta) \otimes \gamma = \alpha \otimes (\beta \otimes \gamma)$). The identity potential ι_r for r has the property that given any potential α for $s \supseteq r$, $\alpha \otimes \iota_r = \alpha$.

Marginalization. The definition of marginalization of potentials (discrete or continuous) depends on the variable being marginalized. Suppose χ is a potential (discrete or continuous) for c, and suppose D is a discrete variable in c. Then the *marginal* of χ by removing D, denoted by χ^{-D}, is the potential for $c \backslash \{D\}$ obtained from χ by addition over the states of D, i.e., $\chi^{-D}(\boldsymbol{x}) = \sum \{\chi(d, \boldsymbol{x}) \mid d \in \Omega_D\}$ for all $\boldsymbol{x} \in \Omega_{c \backslash \{D\}}$.

Suppose χ is a potential (discrete or continuous) for c and suppose X is a continuous variable in c. Then the *marginal* of χ by removing X, denoted by χ^{-X}, is the potential for $c \backslash \{X\}$ obtained from χ by integration over the states of X, i.e., $\chi^{-X}(\boldsymbol{y}) = \int \chi(x, \boldsymbol{y})\, dx$ for all $\boldsymbol{y} \in \Omega_{c \backslash \{X\}}$. If χ contains no equations in X, then the integral is the

usual Riemann integral, and integration is done over Ω_X. If χ contains an equation in X, then the integral is the generalized Riemann integral (also called Riemann-Stieltjes integral), which is defined as follows. First, we solve for X using one of the equations, substitute the solution for X in the other equations and functions (if any), discard the equation used for solving for X, and use the Jacobian as a normalization term when densities are involved. Some examples of generalized Riemann integration are as follows.

$\int [X = c](x)\,(m)\,dx = 1\,(m)$.

$\int [Y = g(X),\ Z = h(X)](x,\,y,\,z)\,(m)\,dx = [Z = h(g^{-1}(Y))](y,\,z)\,(m)$, assuming g is invertible on Ω_X.

$\int [X = c](x)\,f(x)\,dx = f(c)\,(d)$, assuming f is a density function.

$\int [Y = g(X)](x,\,y)\,f(x)\,dx = |(d/dy)(g^{-1}(y))|\,f(g^{-1}(y))\,(d)$, assuming f is a density function, and g is invertible and differentiable on Ω_X. $|(d/dy)(g^{-1}(y))|$ is called the *Jacobian*.

$\int [Y = g(X),\ Z = h(X)](x,\,y,\,z)\,f(x)\,dx = [Z = h(g^{-1}(Y))](y,\,z)\,|(d/dy)(g^{-1}(y))|\,f(g^{-1}(y))\,(d)$, assuming f is a density function, and g is invertible and differentiable on Ω_X.

In a Bayesian network, each variable X is associated with a conditional probability function for X given its parents, $pa(X)$. This conditional probability function is represented by a potential for $\{X\} \cup pa(X)$ called the conditional associated with X.

If α is a conditional potential associated with A, and its domain is a (i.e., $a = \{A\}\cup pa(X)$), then α^{-A} is an identity potential for $a\backslash\{A\} = pa(X)$, the parents of A, i.e., if β is any potential whose domain contains $a\backslash\{A\}$, then $\alpha^{-A}\otimes\beta = \beta$.

To reverse an arc (X, Y) in a Bayesian network, we compute the marginal $(\xi\otimes\psi)^{-X}$, where ξ is the potential associated with X (representing the conditional for X given $pa(X)$), and ψ is the potential associated with Y (representing the conditional for Y given $pa(Y)$). The potential $(\xi\otimes\psi)^{-X}$ represents the conditional for Y given $pa(X)\cup pa(Y)\backslash\{X\}$, and its nature (discrete or continuous) depends on Y. Thus, if Y is discrete, then $(\xi\otimes\psi)^{-X}$ is a discrete potential, and if Y is continuous or deterministic, then $(\xi\otimes\psi)^{-X}$ is a continuous potential. Furthermore, if ξ and ψ are both equation potentials, or if ξ is discrete and ψ is an equation potential, then $(\xi\otimes\psi)^{-X}$ is an equation potential. In both of these cases, the units of $(\xi\otimes\psi)^{-X}$ are in probability mass. In all other cases (see Table 2.1), $(\xi\otimes\psi)^{-X}$ is a density potential.

Table 2.1. The nature of the $(\xi \otimes \psi)^{-X}$ potential associated with Y

If ξ assoc. with X is:	and ψ assoc. with Y is:	then $(\xi \otimes \psi)^{-X}$ assoc. with Y is:
discrete, density, or equation	discrete	discrete
equation	equation	equation
discrete	equation	equation
density	density	density
density	equation	density
equation	density	density
discrete	density	density

Divisions. Arc reversals involve divisions of potentials, and the potential in the denominator is always a marginal of the potential in the numerator. Suppose (X, Y) is a reversible arc in a Bayesian network, suppose ξ is a potential for $\{X\} \cup pa(X)$ associated with X, and suppose ψ is a potential for $\{Y\} \cup pa(Y)$ associated with Y. After reversing the arc (X, Y), the revised potential associated with X is $(\xi \otimes \psi) \% (\xi \otimes \psi)^{-X}$. The definition of $(\xi \otimes \psi) \% (\xi \otimes \psi)^{-X}$ is as follows. If ξ is a density potential (for $\{X\} \cup pa(X)$) and ψ is an equation potential (for $\{Y\} \cup pa(Y)$), then $(\xi \otimes \psi) \% (\xi \otimes \psi)^{-X} = \psi$. In all other cases, $(\xi \otimes \psi) \% (\xi \otimes \psi)^{-X}$ is a potential for $\{Y\} \cup pa(X) \cup pa(Y)$ obtained from $(\xi \otimes \psi)$ and $(\xi \otimes \psi)^{-X}$ by point-wise division, i.e., $((\xi \otimes \psi) \% (\xi \otimes \psi)^{-X})(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = (\xi \otimes \psi)(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) / ((\xi \otimes \psi)^{-X})(y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t})$ for all $x \in \Omega_X$, $y \in \Omega_Y$, $\boldsymbol{r} \in \Omega_{pa(X) \setminus pa(Y)}$, $\boldsymbol{s} \in \Omega_{pa(X) \cap pa(Y)}$, $\boldsymbol{t} \in \Omega_{pa(Y) \setminus (\{X\} \cup pa(X))}$. Notice that if $((\xi \otimes \psi)^{-X})(y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = 0$, then $(\xi \otimes \psi)(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = 0$. In this case, we will simply define 0/0 as 0.

The quotient $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ represents the conditional for X given $pa(X)\cup pa(Y)\cup\{Y\}$, and its nature depends on X. Thus, if X is discrete, then $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ is a discrete potential, and if X is continuous or deterministic, then $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ is a continuous potential. More specifically, the nature of $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ is as described in Table 2.2.

Table 2.2: The nature of the $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ potential associated with X

If ξ assoc. with X is:	and ψ assoc. with Y is:	then $(\xi\otimes\psi)^{-X}$ assoc. with Y is:	and $(\xi\otimes\psi)\%(\xi\otimes\psi)^{-X}$ assoc. with X is:
discrete	discrete	discrete	discrete
discrete	density	density	discrete
discrete	equation	equation	discrete
density	discrete	discrete	density
density	density	density	density
density	equation	density	equation
equation	discrete	discrete	equation
equation	density	density	equation
equation	equation	equation	equation

The rationale for the separate definition of division for the case where ξ is a density potential and ψ is an equation potential is as follows. Consider the Bayesian network shown in Figure 2.3 consisting of two continuous variables X and Y, where X has PDF $f(x)$, and Y is a deterministic function of X, say $Y|x = g(x)$ with probability 1, where g is invertible and differentiable in Ω_x. Let ξ and ψ denote the density and equation potentials associated with X and Y, respectively. Then $\xi(x) = f(x)$, and $\psi(x, y) = [Y = g(X)](x, y)$. After reversal of the arc (X, Y), the revised potential associated

with Y is $\psi'(y) = (\xi \otimes \psi)^{-X}(y) = \int f(x) \, [Y = g(X)](x, y) \, dx = |(d/dy)(g^{-1}(y))| \, f(g^{-1}(y))$. The Jacobian term $|(d/dy)(g^{-1}(y))|$ in the potential ψ' is a consequence of expressing the conditional for X in units of probability density, and the conditional for Y in units of probability mass. If we had expressed the conditional for X in units of probability mass by using the cumulative distribution function (CDF) $F(x)$, then a Jacobian term would not have been required—the CDF of Y is given by $P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F(g^{-1}(y))$. After arc reversal, the revised potential associated with X is $\xi' = (\xi \otimes \psi)\%(\xi \otimes \psi)^{-X}$. The numerator of this expression, $\xi \otimes \psi$, does not have any probabilistic semantics since ξ has units of probability density and ψ has units of probability mass. However, if we define the division of potentials so that the Jacobian term in the denominator disappears, we would obtain the correct results. Thus,

$$\xi'(x, y) = (\xi \otimes \psi)(x, y)/(\xi \otimes \psi)^{-X}(y) = f(x) \, [Y = g(X)](x, y) \, / \, (|(d/dy)(g^{-1}(y))| \, f(g^{-1}(y)))$$

$$= [Y = g(X)](x, y) = [X = g^{-1}(Y)](x, y),$$

which is an equation potential. Also, we would have to ignore the Jacobian term in the combination so that $\xi \otimes \psi = \xi' \otimes \psi'$, i.e., $f(x) \, [Y = g(X)](x, y) = [X = g^{-1}(Y)](x, y) \, |(d/dy)(g^{-1}(y))| \, f(g^{-1}(y))$.
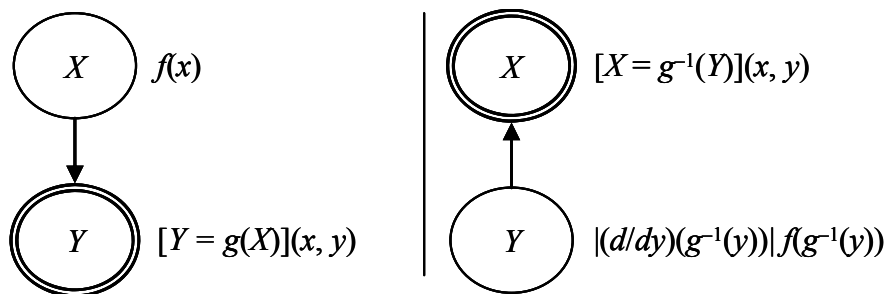
Figure 2.3. Arc reversal between a continuous and a deterministic variable.

2.2.2   Arc Reversals

This section describes arc reversals between every possible kind of pairs of variables. Given a Bayesian network graph, i.e., a directed acyclic graph, there always exists a sequence of variables such that whenever there is an arc (X, Y) in the network, X precedes Y in the sequence. An arc (X, Y) can be reversed only if there exists a sequence such that X and Y are adjacent in this sequence.

In a Bayesian network, each variable is associated with a conditional potential representing the conditional distribution for it given its parents. A fundamental assumption of Bayesian network theory is that the combination of all the conditional potentials is the joint distribution of all variables in the network. Suppose (X, Y) is an arc in a Bayesian network such that X and Y are adjacent, and suppose ξ and ψ are the potentials associated with X and Y, respectively. Let $pa(X) = r \cup s$, and $pa(Y) = s \cup t$. Since X and Y are adjacent, the variables in $r \cup s \cup t$ precede X and Y in a sequence compatible with the arcs. Then $\xi \otimes \psi$ represents the conditional joint distributions of $\{X, Y\}$ given $r \cup s \cup t$, $(\xi \otimes \psi)^{-X}$ represents the conditional distributions of Y given $r \cup s \cup t$, and $(\xi \otimes \psi)\%(\xi \otimes \psi)^{-X}$ represents the conditional distributions of X given $r \cup s \cup t \cup \{Y\}$. If the arc (X, Y) is reversed, the potentials ξ and ψ associated with X and Y are replaced by potentials $\xi' = (\xi \otimes \psi)\%(\xi \otimes \psi)^{-X}$, and $\psi' = (\xi \otimes \psi)^{-X}$, respectively. This general case is illustrated in Figure 2.4. Although X and Y are shown as continuous nodes, they can each be discrete or deterministic.
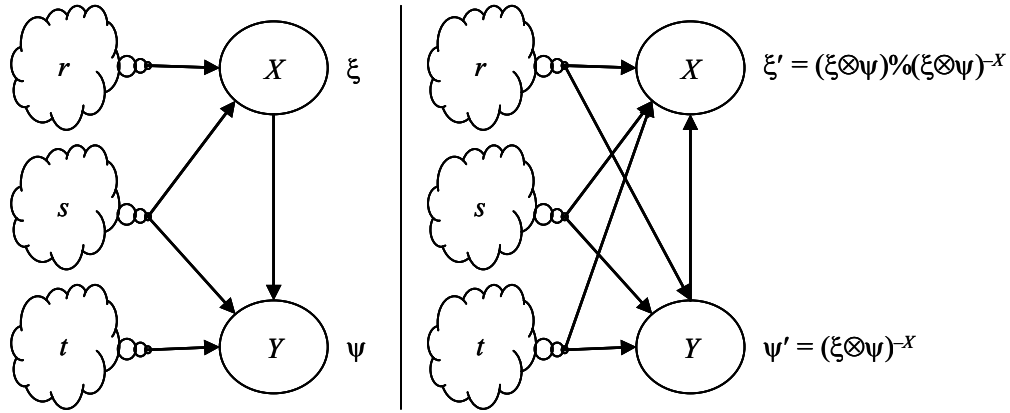
Figure 2.4. Reversal of arc (X, Y)

Some observations about the arc reversal process are as follows: First, arc reversal is a local operation that affects only the potentials associated with the two variables defining the arc. The potentials associated with the other variables remain unchanged.

Second, notice that $\xi \otimes \psi = \xi' \otimes \psi'$. Thus, the joint conditional distributions of $\{X, Y\}$ given $r \cup s \cup t$ remain unchanged by arc reversal. Also, since the other potentials for $r \cup s \cup t$ do not change, the joint distribution of all variables in a Bayesian network remains unchanged.

Third, for any potential α, let $dom(\alpha)$ denote the domain of α. Notice that the $dom(\xi') = dom(\xi) \cup dom(\psi) = r \cup s \cup t \cup \{X\} \cup \{Y\}$, and the $dom(\psi') = dom(\xi) \cup dom(\psi) \backslash \{X\} = r \cup s \cup t \cup \{Y\}$. Thus after arc reversal, X and Y inherit each other's parents, Y loses X as a parent, and X gains Y as a parent.

Fourth, suppose we reverse the arc (Y, X) in the revised Bayesian network. Let ξ'' and ψ'' denote the potentials associated with X and Y after reversal of arc (Y, X). Then

$$\xi'' = (\xi' \otimes \psi')^{-Y} = (((\xi \otimes \psi)\%(\xi \otimes \psi)^{-X}) \otimes (\xi \otimes \psi)^{-X})^{-Y} = (\xi \otimes \psi)^{-Y} = \xi \otimes (\psi^{-Y}) = \xi \otimes \iota_{pa(Y)},$$

and

$$\psi'' = (\xi'\otimes\psi')\%(\xi'\otimes\psi')^{-Y} = (\xi\otimes\psi)\%(\xi\otimes\iota_{pa(Y)}) = \psi\otimes\iota_{\{X\}\cup pa(X)}.$$

If we ignore the identity potentials (since these have no effect on the joint distribution), ξ'' and ψ'' are the same as ξ and ψ, what we started with.

2.2.2.1   Discrete to Discrete

In this section we describe reversal of an arc between two discrete nodes. This is the standard case and we discuss it here only for completeness.

Consider the BN given on the left-hand side of Figure 2.5. Let δ and ε denote the discrete potentials associated with variables D and E, respectively, before arc reversal, and δ' and ε' after arc reversal. Then, for all $e_j \in \Omega_E$, and $d_i \in \Omega_D$,

$\delta(a,\, b,\, d_i) = P(d_i|a,\, b),$

$\varepsilon(b,\, c,\, d_i,\, e_j) = P(e_j|b,\, c,\, d_i),$

$\varepsilon'(a,\, b,\, c,\, e_j) = (\delta\otimes\varepsilon)^{-D}(a,\, b,\, c,\, e_j) = \Sigma\{P(d_i|a,\, b)\, P(e_j|b,\, c,\, d_i) \mid d_i \in \Omega_D\},$ and

$\delta'(a,\, b,\, c,\, d_i,\, e_j) = ((\delta\otimes\varepsilon)\%(\delta\otimes\varepsilon)^{-D})(a,\, b,\, c,\, d_i,\, e_j)$

$$= P(d_i|a,\, b)\, P(e_j|b,\, c,\, d_i) \,/\, \Sigma\{P(d_i|a,\, b)\, P(e_j|b,\, c,\, d_i) \mid d_i \in \Omega_D\}.$$

The resulting BN is given on the right-hand side of Figure 2.5.
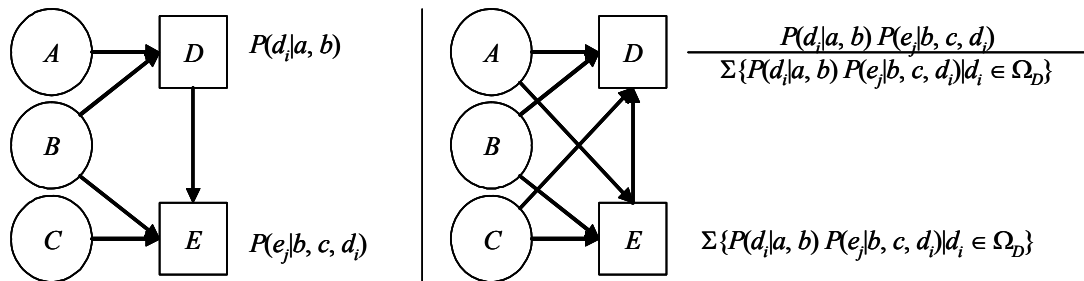
Figure 2.5. Arc reversal between two discrete nodes.

2.2.2.2 Continuous to Continuous

In this section, we describe arc reversals between two continuous variables. Consider the BN given on the left-hand side of Figure 2.6. In this BN, X has conditional PDF $f(u, v, x)$ and Y has conditional PDF $g(v, w, x, y)$. Let ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then,

$$\xi(u, v, x) = f(u, v, x),$$

$$\psi(v, w, x, y) = g(v, w, x, y),$$

$$\psi'(u, v, w, y) = (\xi \otimes \psi)^{-X}(u, v, w, y) = \int f(u, v, x)\, g(v, w, x, y)\, dx,$$

$$\xi'(u, v, w, x, y) = ((\xi \otimes \psi)\%(\xi \otimes \psi)^{-X})(u, v, w, x, y)$$

$$= f(u, v, x)\, g(v, w, x, y) \, / \, \left(\int f(u, v, x)\, g(v, w, x, y)\, dx\right).$$

The resulting BN is shown on the right-hand side of Figure 2.6.
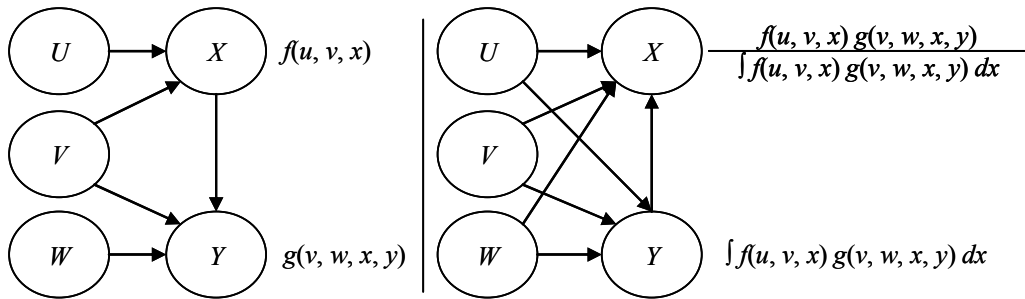

Figure 2.6. Arc reversal between two continuous nodes.

2.2.2.3 Continuous to Deterministic

As we have already discussed, the arc reversal between a continuous and a deterministic variable is slightly different from the arc reversal between two continuous variables since their joint probability density function does not exist. The

conditional associated with the continuous variable has units of density, and the conditional associated with the deterministic variable has units of mass. After arc reversal, we transfer the density from the continuous node to the deterministic node, which results in the deterministic node being continuous and the continuous node being deterministic.

Consider the situation shown in Figure 2.7. In this BN, X has continuous parents U and V, and Y has continuous parents V and W in addition to X. The density at X is f and the equation at Y is $[Y = h(V, W, X)]$. We assume h is invertible in X and differentiable on Ω_X. The potentials before and after arc reversals are as follows.

$$\xi(u, v, x) = f(u, v, x),$$

$$\psi(v, w, x, y) = [Y = h(V, W, X)](v, w, x, y),$$

$$\psi'(u, v, w, y) = (\xi \otimes \psi)^{-X}(u, v, w, y) = \int f(u, v, x) \, [Y = h(V, W, X)](v, w, x, y) \, dx$$

$$= |(\partial/\partial y)(h^{-1}(v, w, y))| f(u, v, h^{-1}(v, w, y)), \text{ and}$$

$$\xi'(u, v, w, x, y) = ((\xi \otimes \psi)\%(\xi \otimes \psi)^{-X})(u, v, w, x, y) = \psi(v, w, x, y)$$

$$= [Y = h(V, W, X)](v, w, x, y) = [X = h^{-1}(V, W, Y)](v, w, x, y).$$

After we reverse the arc (X, Y), both X and Y inherit each other's parents, but X loses U as a parent. Also, Y has a density function and X has a deterministic conditional distribution. The resulting BN is given on right-hand side of Figure 2.7. Some of the qualitative conclusions here, namely X loses U as a parent, Y has a density function, and X has a deterministic conditional distribution, are based on the assumption that U, V, W are continuous. If any of these are discrete, the conclusion can change as we will demonstrate in Section 2.2.3.
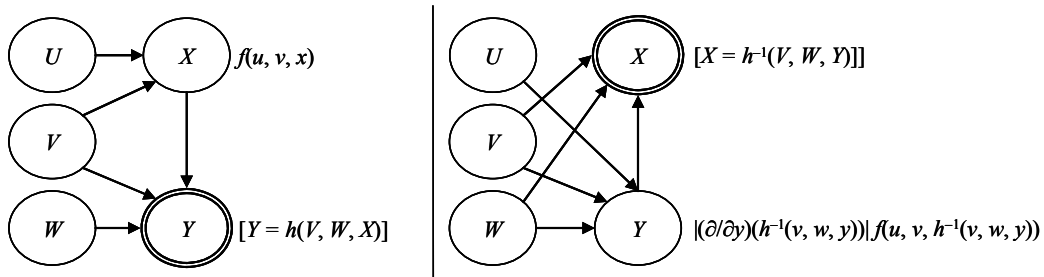
Figure 2.7. Arc reversal between a continuous and a deterministic variable.

As an example of the general case, consider the Bayesian network consisting of two continuous variables and a deterministic variable whose function is the sum of its two parents as shown in Figure 2.8. $X \sim f(x)$, $Y|x \sim g(x, y)$, and $Z = X + Y$. Let ξ, ψ, ζ denote the potentials associated with X, Y, and Z, respectively, before arc reversal, and ψ' and ζ' denote the revised potentials associated with Y and Z, respectively, after reversal of arc (Y, Z). Then,

$\xi(x) = f(x)$,

$\psi(x, y) = g(x, y)$,

$\zeta(x, y, z) = [Z = X + Y](x, y, z)$,

$\zeta'(x, z) = (\psi \otimes \zeta)^{-Y}(x, z) = \int g(x, y) [Z = X + Y](x, y, z) \, dy = g(x, z - x)$, and

$\psi'(x, y, z) = ((\psi \otimes \zeta)\%(\psi \otimes \zeta)^{-Y})(x, y, z) = \zeta(x, y, z) = [Z = X + Y](x, y, z)$

$= [Y = Z - X](x, y, z)$.

If we reverse the arc (X, Z) in the revised Bayesian network, we obtain the marginal distribution of Z,

$\zeta''(z) = (\xi \otimes \zeta')^{-X}(z) = \int f(x) \, g(x, z - x) \, dx$,

which is the convolution formula for Z. The revised potential at X,

$$\xi'(x, z) = ((\xi \otimes \zeta')\% (\xi \otimes \zeta')^{-X})(x, z) = f(x)\, g(x, z - x)/(\int f(x)\, g(x, z - x)\, dx),$$

represents the conditional distribution of X given z.
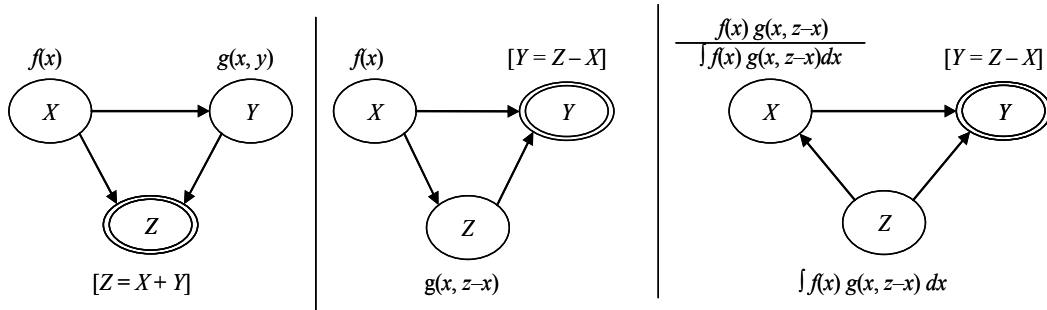
Figure 2.8. A continuous Bayesian network with a deterministic variable.

We have assumed that the function describing the deterministic variable is invertible and differentiable. Let us consider the case where the function is not invertible, but it is "piecewise invertible," i.e., invertible in some known deterministic regions of Ω_X. For example, consider a Bayesian network with two continuous variables X and Y, where X has PDF $f(x)$ and Y is a deterministic function of X described by the function $Y = X^2$ as shown in Figure 2.9.
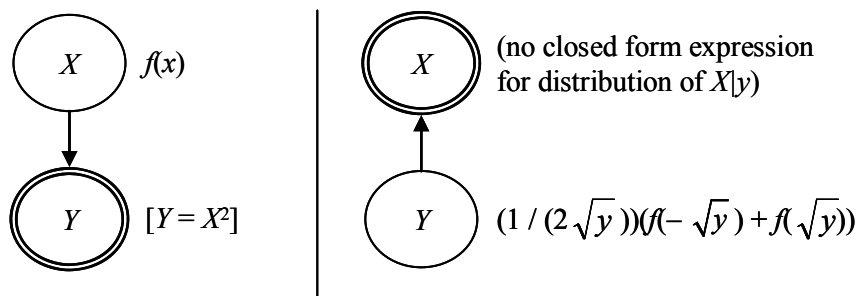
Figure 2.9. A continuous Bayesian network with a piecewise invertible deterministic variable.

This function is not invertible, but is piecewise invertible since it is invertible in the regions $(-\infty, 0]$ and $(0, \infty)$. Thus, the deterministic function can be written as

follows: $[Y = X^2](x, y) = [Y = X^2, X \le 0](x, y) + [Y = X^2, X > 0](x, y)$. Here, $[Y = X^2, X \le 0]$ and $[Y = X^2, X > 0]$ are indicator functions in the usual sense, i.e., $[Y = X^2, X \le 0](x, y) = 1$ if $y = x^2$ and $x \le 0$, and $= 0$ otherwise. In this case, we can compute the marginal for Y in closed form, but not the conditional for X given Y. Suppose ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then

$\xi(x) = f(x)$,

$\psi(x, y) = [Y = X^2](x, y)$,

$\psi'(y) = (\xi \otimes \psi)^{-X}(y) = \int f(x) \, [Y = X^2](x, y) \, dx$

$\qquad = \int f(x) \, [Y = X^2](x, y) \, ([X \le 0](x) + [X > 0](x)) \, dx$

$\qquad = \int f(x) \, [Y = X^2, X \le 0](x, y) \, dx + \int f(x) \, [Y = X^2, X > 0](x, y) \, dx$

$\qquad = (1/(2\sqrt{y})) f(-\sqrt{y}) \, [X \le 0](-\sqrt{y}) + (1/(2\sqrt{y})) f(\sqrt{y}) \, [X > 0](\sqrt{y})$

$\qquad = (1/(2\sqrt{y})) (f(-\sqrt{y}) + f(\sqrt{y}))$, for all $y > 0$.

Since the deterministic function is not invertible, there is no closed form expression for $\xi' = (\xi \otimes \psi) \% (\xi \otimes \psi)^{-X}$.

If a deterministic variable has a function that is neither invertible nor piecewise invertible, then we cannot even describe the potential associated with Y in closed form. One example of such a function is $Z = Max\{X, Y\}$. In this case, if we wish to compute the marginal of Z, we can approximate this function, e.g., with a mixture of Gaussians [Shenoy 2006]. This will be illustrated in Section 2.3.2.

2.2.2.4 Deterministic to Continuous

In this subsection, we describe arc reversal between a deterministic and a continuous variable. Consider a BN as shown on the left-hand side of Figure 2.10. X is a deterministic variable associated with a function, $[X = h(U, V)]$, and Y is a continuous variable and the conditional distribution of $Y|(v, w, x)$ is distributed as $g(v, w, x, y)$. Suppose we wish to reverse the arc (X, Y). Since there is no density potential at X, Shenoy [2006] suggests to first reverse arc (U, X) or (V, X) (resulting in a density potential at X), and then to reverse arc (X, Y) using the rules for arc reversal between two continuous nodes. However, here we show that it is possible to reverse an arc between a deterministic node and a continuous node directly without having to reverse other arcs.
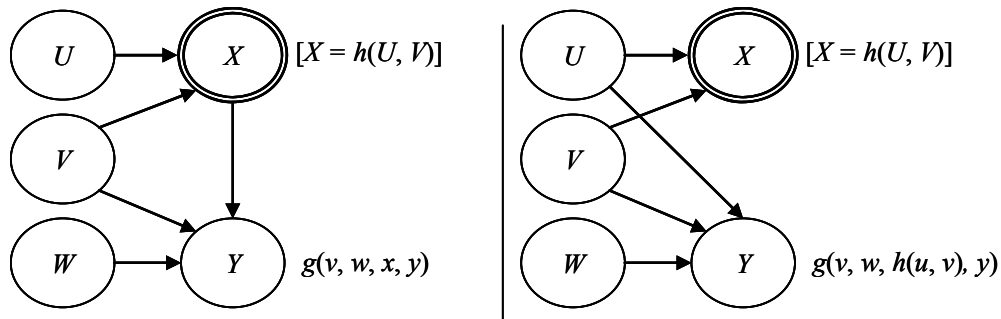
Figure 2.10. Arc reversal between a deterministic and a continuous node.

Consider again the BN given on left-hand side of Figure 2.10. Suppose we wish to reverse the arc (X, Y). Let ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then,

$\xi(u, v, x) = [X = h(U, V)](u, v, x),$

$\psi(v, w, x, y) = g(v, w, x, y),$

$$\psi'(u,\ v,\ w,\ y)\ =\ ((\xi\otimes\psi)^{-X})(u,\ v,\ w,\ y)\ =\ \int [X=h(U,V)](u,v,x)\ g(v,w,x,y)\ dx$$

$$=\ g(v,\ w,\ h(u,v),\ y),\ \text{and}$$

$$\xi'(u,v,w,x,y)\ =\ (\xi\otimes\psi)\%((\xi\otimes\psi)^{-X})(u,v,\quad w,\quad x,\quad y)\ =\ \xi(u,\quad v,\quad x)$$

$$=\ [X=h(U,V)](u,v,x).$$

Notice that, that ξ' does not depend on either W or Y. Thus, after arc reversal, there is no arc from Y to X, i.e., the arc being reversed disappears, and X does not inherit an arc from W. The resulting BN is shown on the right-hand side of Figure 2.10.

2.2.2.5 Deterministic to Deterministic

In this subsection, we describe arc reversal between two deterministic variables. Consider the BN on the left-hand side of Figure 2.11. X is a deterministic function of its parents $\{U,\ V\}$, and Y is also a deterministic function of its parents $\{X,\ V,\ W\}$. Suppose we wish to reverse the arc $(X,\ Y)$. Let ξ and ψ denote the potentials associated with X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then
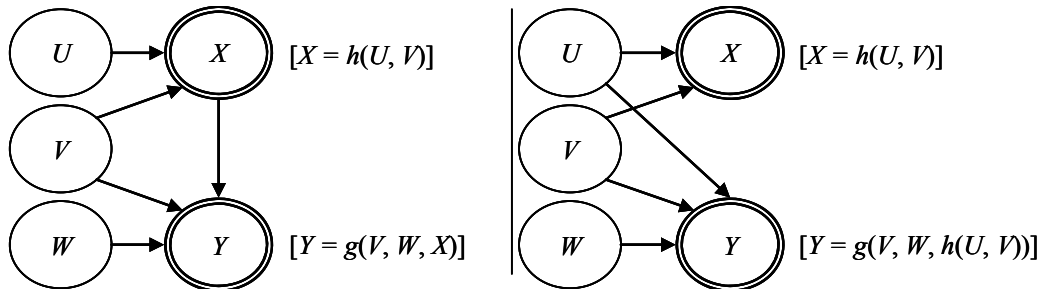
Figure 2.11. Arc reversal between two deterministic nodes.

$$\xi(u,\ v,\ x)=[X=h(U,V)](u,v,x),$$

$$\psi(v, w, x, y) = [Y = g(V, W, X)](v, w, x, y),$$

$$\psi'(u, v, w, y) = (\xi \otimes \psi)^{-X}(u, v, w, y)$$

$$= \int [X = h(U, V)](u, v, x) \, [Y = g(V, W, X)](v, w, x, y) \, dx$$

$$= [Y = g(V, W, h(U, V))](u, v, w, y), \text{ and}$$

$$\xi'(u, v, w, x, y) = ((\xi \otimes \psi)\%(\xi \otimes \psi)^{-X})(u, v, w, x, y)$$

$$= [X = h(U, V)](u, v, x) \, [Y = g(V, W, X)](v, w, x, y) \, / \, [Y = g(V, W, h(U, V))]$$

$$(u, v, w, y) = [X = h(U, V)](u, v, x).$$

Notice that ξ' does not depend on either Y nor W. The arc being reversed disappears, and X does not inherit a parent of Y.

2.2.2.6   Continuous to Discrete

In this section, we will describe arc reversal between a continuous and a discrete node. Consider the Bayesian network as shown in Figure 2.12. X is a continuous node with conditional PDF $f(u, v, x)$, and D is a discrete node with conditional masses $P(d_i|v, w, x)$ for each $d_i \in \Omega_D$. Let ξ and δ denote the density and discrete potentials associated with X and D, respectively, before arc reversal, and ξ' and δ' after arc reversal. Then

$$\xi(u, v, x) = f(u, v, x),$$

$$\delta(v, w, x, d_i) = P(d_i|v, w, x),$$

$$\delta'(u, v, w, d_i) = (\xi \otimes \delta)^{-X}(u, v, w, d_i) = \int f(u, v, x) \, P(d_i|v, w, x) \, dx, \text{ and}$$

$$\xi'(u, v, w, x, d_i) = ((\xi \otimes \delta)\%(\xi \otimes \delta)^{-X})(u, v, w, x, y)$$

$$= f(u, v, x) \, P(d_i|v, w, x) \, / \, (\int f(u, v, x) \, P(d_i|v, w, x) \, dx).$$

The BN on the RHS of Figure 2.12 depicts the results after arc reversal.
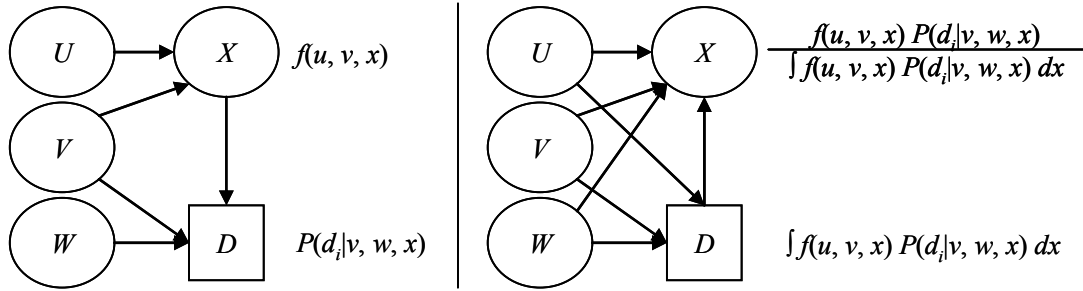
Figure 2.12. Arc reversal between a continuous and a discrete node.

For a concrete example, consider the simpler hybrid BN shown on the LHS of Figure 2.13. X is a continuous variable, distributed as $N(0, 1)$. D is a discrete variable with two states $\{d_1, d_2\}$. The conditional probability mass functions of D are as follows: $P(d_1|x) = 1/(1 + e^{-2x})$ and $P(d_2|x) = e^{-2x}/(1 + e^{-2x})$. Let δ and ξ denote the potentials associated with D and X, respectively, before arc reversal, and δ' and ξ' after arc reversal. Then,

$$\delta(d_1, x) = 1/(1 + e^{-2x}), \delta(d_2, x) = e^{-2x}/(1 + e^{-2x}),$$

$$\xi(x) = \varphi_{0, 1}(x),$$

$$\delta'(d_1) = (\delta \otimes \xi)^{-X}(d_1) = \int (1/(1 + e^{-2x})) \, \varphi_{0, 1}(x) \, dx = 0.5,$$

$$\delta'(d_2) = (\delta \otimes \xi)^{-X}(d_2) = \int (e^{-2x}/(1 + e^{-2x}) \, \varphi_{0, 1}(x) dx = 0.5,$$

$$\xi'(d_1, x) = ((\delta \otimes \xi)\%(\delta \otimes \xi)^{-X})(d_1, x) = (1/(1 + e^{-2x})) \, \varphi_{0, 1}(x))/0.5$$

$$= (2/(1 + e^{-2x})) \, \varphi_{0, 1}(x),$$

$$\xi'(d_2, x) = (\delta \otimes \xi)\%(\delta \otimes \xi)^{-X}(d_2, x) = (e^{-2x}/(1 + e^{-2x})) \, \varphi_{0, 1}(x))/0.5$$

$$= (2e^{-2x}/(1 + e^{-2x})) \, \varphi_{0, 1}(x),$$

The resulting BN after the arc reversal is given on the RHS of Figure 2.13.
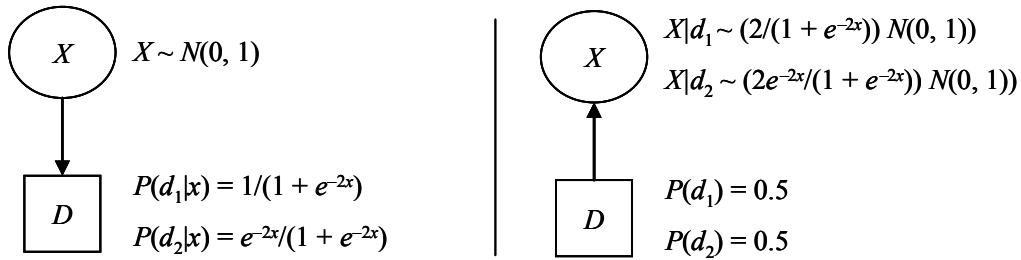
Figure 2.13. Arc reversal between a continuous and a discrete node.

2.2.2.7   Deterministic to Discrete

In this subsection, we describe reversal of an arc between a deterministic and a discrete variable. Consider the hybrid BN shown on the left-hand side of Figure 2.14. Let ξ and δ denote the potentials at X and D, respectively, before arc reversal, and let ξ' and δ' denote the potentials after arc reversal. Then,

$$\xi(u, v, x) = [X = h(U, V)](u, v, x),$$

$$\delta(v, w, x, d_i) = P(d_i|v, w, x),$$

$$\delta'(u, v, w, d_i) = \int [X = h(U, V)](u, v, x)\, P(d_i|v, w, x)\, dx = P(d_i|v, w, h(u, v)), \text{ and}$$

$$\xi'(u, v, w, x, y) = [X = h(U, V)](u, v, x)\, P(d_i|v, w, x)/P(d_i|v, w, h(u, v))$$

$$= [X = h(U, V)](u, v, x).$$
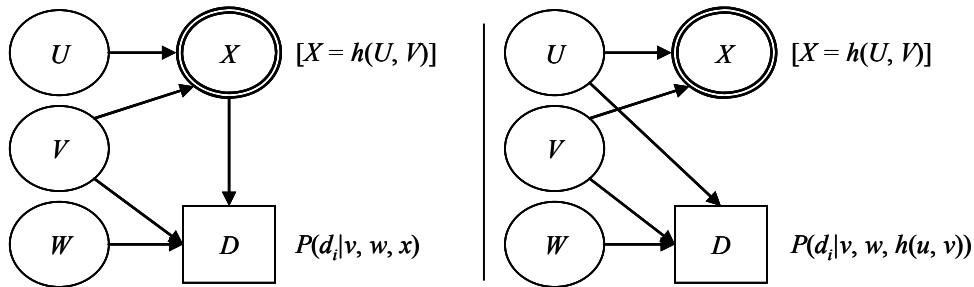
Figure 2.14. Arc reversal between a deterministic and a discrete variable.

Notice that ξ′ depends on neither D nor W. The illustration of an arc reversal between a deterministic and discrete node with parents is given in Figure 2.14 above.

2.2.2.8   Discrete to Continuous

In this subsection, we describe reversal of an arc from a discrete to a continuous variable. Consider the hybrid BN shown on the LHS of Figure 2.15. Let δ and ξ denote the potentials associated with D and X, respectively, before arc reversal, and δ′ and ξ′ after arc reversal. Then,

$$\delta(u,\, v,\, d_i) = P(d_i|u,\, v),$$

$$\xi(v,\, w,\, x,\, d_i) = f(v,\, w,\, x,\, d_i),$$

$$\xi'(u,\, v,\, w,\, x) = (\delta \otimes \xi)^{-D}(u,\, v,\, w,\, x) = \Sigma\{P(d_i|u,\, v)\, f(v,\, w,\, x,\, d_i) \mid d_i \in \Omega_D\},$$

$$\delta'(u,\, v,\, w,\, x,\, d_i) = ((\delta \otimes \xi)\%(\delta \otimes \xi)^{-D})(u,\, v,\, w,\, x,\, d_i)$$

$$= P(d_i|u,\, v)\, f(v,\, w,\, x,\, d_i)/\Sigma\{P(d_i|u,\, v)\, f(v,\, w,\, x,\, d_i) \mid d_i \in \Omega_D\}.$$

The density at X after arc reversal is a mixture density.
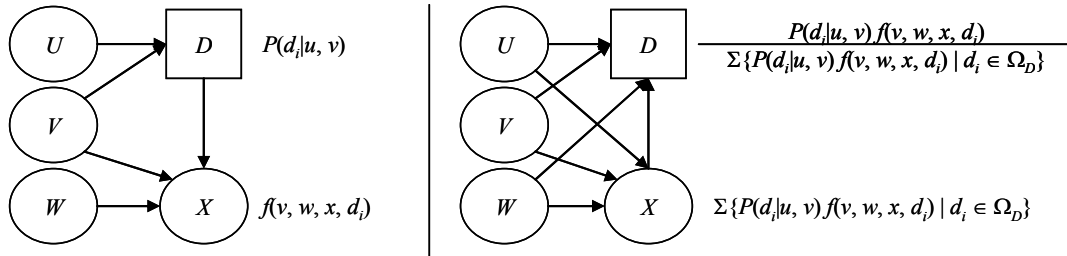
Figure 2.15. Arc reversal between a discrete and a continuous variable.

For a concrete example, consider the BN given on the LHS of Figure 2.16. The discrete variable D has two states $\{d_1,\, d_2\}$ with $P(d_1) = 0.5$ and $P(d_2) = 0.5$. X is a continuous variable whose conditional distributions are $X|d_1 \sim N(0,\, 1)$ and

$X|d_2 \sim N(2, 1)$. Let δ and ξ denote the potentials associated with X and D, respectively, before arc reversal, and δ' and ξ' after arc reversal. Then,

$$\delta(d_1) = 0.5, \delta(d_2) = 0.5,$$

$$\xi(d_1, x) = \varphi_{0,1}(x), \xi(d_2, x) = \varphi_{2,1}(x),$$

$$\xi'(x) = (\delta \otimes \xi)^{-D}(x) = 0.5 \, \varphi_{0,1}(x) + 0.5 \, \varphi_{2,1}(x),$$

$$\delta'(d_1, x) = ((\delta \otimes \xi)\%(\delta \otimes \xi)^{-D})(d_1, x) = (0.5 \, \varphi_{0,1}(x))/(0.5 \, \varphi_{0,1}(x) + 0.5 \, \varphi_{2,1}(x)),$$

$$\delta'(d_2, x) = ((\delta \otimes \xi)\%(\delta \otimes \xi)^{-D})(d_2, x) = (0.5 \, \varphi_{2,1}(x))/(0.5 \, \varphi_{0,1}(x) + 0.5 \, \varphi_{2,1}(x)),$$

The resulting BN after the arc reversal is given on the RHS of Figure 2.16.
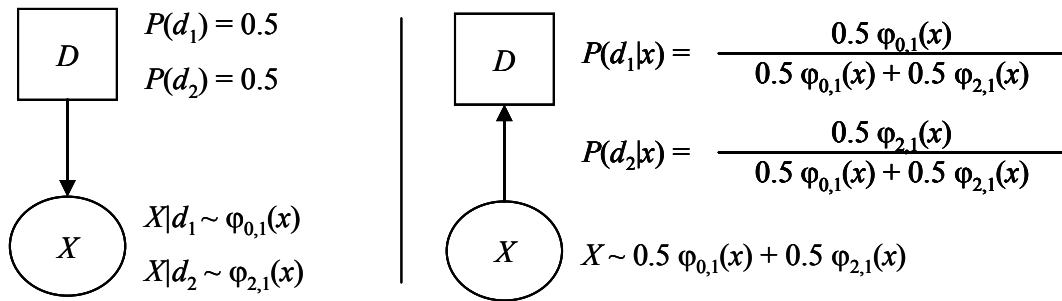
Figure 2.16. An example of an arc reversal between a discrete and a continuous variable.

2.2.2.9 Discrete to Deterministic

In this subsection, we describe reversal of an arc between a discrete and a deterministic variable. Consider the hybrid BN as shown on the left-hand side of Figure 2.17. Suppose that $\Omega_D = \{d_1, \ldots, d_k\}$. Let δ and ξ denote the potentials associated with D and X, respectively, before arc reversal, and δ' and ξ' after arc reversal. Then,
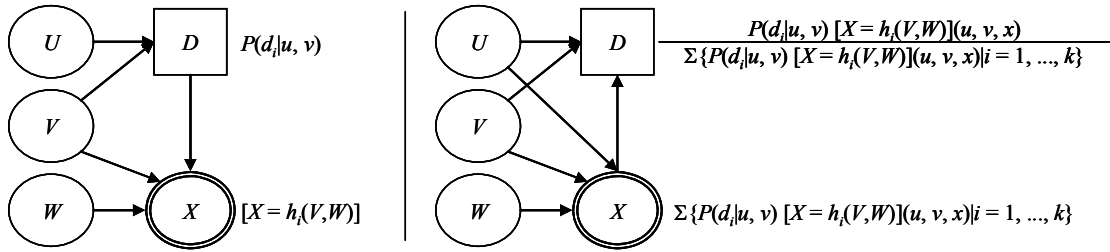
Figure 2.17. Arc reversal between a discrete and a deterministic variable.

$$\delta(u, v, d_i) = P(d_i|u, v),$$

$$\xi(v, w, x, d_i) = [X = h_i(V, W)](v, w, x),$$

$$\xi'(u, v, w, x) = (\delta \otimes \xi)^{-D}(u, v, w, x) = \Sigma\{P(d_i|u, v) [X = h_i(V, W)](v, w, x) \mid i = 1, \ldots,$$

$$k\},$$

$$\delta'(u, v, w, x, d_i) =$$

$$P(d_i|u, v) [X = h_i(V, W)](v, w, x)/\Sigma\{P(d_i|u, v) [X = h_i(V, W)](v, w, x) \mid i = 1,$$

$$\ldots, k\}.$$

The situation after arc reversal is shown on the right-hand side of Figure 2.17. Notice that after arc reversal, X has a weighted set of equation functions. Since the values of ξ' are in units of probability mass, X remains deterministic after arc reversal.

For a concrete example, consider the simpler hybrid BN shown on the LHS of Figure 2.18. V has the uniform distribution on $(0, 1)$. D has two states $\{d_1, d_2\}$ with $P(d_1|v) = [0 < V \le 0.5](v)$, and $P(d_2|v) = [0.5 < V < 1](v)$. Here, $[0 < V \le 0.5]$ and $[0.5 < V < 1]$ are indicator functions in the usual sense. X is deterministic with equation functions $[X = V]$ if $D = d_1$, and $[X = -V]$ if $D = d_2$. After arc reversal, the conditional distributions at D and X are as shown in the RHS of Figure 2.18 (these are

special cases of the general formulae given in Figure 2.17). Let ϖ denote the continuous potential at V. Then $\varpi(v) = [0 < V < 1](v)$. We can find the marginal of X from the BN on the RHS of Figure 2.18 by reversing arc (V, X) as follows.

$$(\varpi \otimes \xi')^{-V}(x) = \int [0 < V < 1](v) \, [0 < V \leq 0.5](v) \, [X = V](x, v) \, dv +$$

$$\int [0 < V < 1](v) \, [0.5 < V < 1](v) \, [X = -V](x, v) \, dv$$

$$= [0 < X \leq 0.5](x) + [-1 < X < -0.5](x).$$

Thus, the marginal distribution of X is uniform on the interval $(-1, -0.5) \cup (0, 0.5)$.
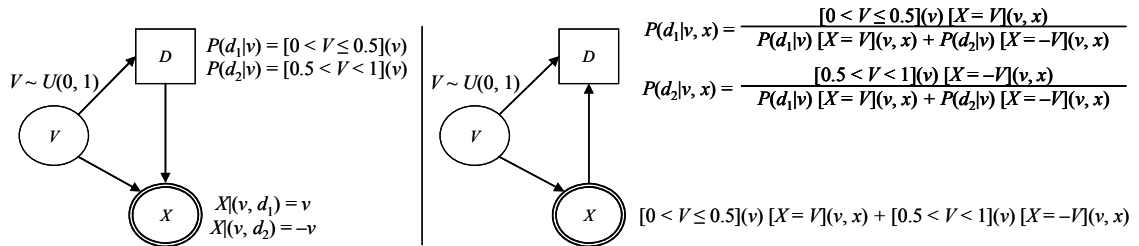
Figure 2.18. A special case of arc reversal between a discrete and deterministic variable.

2.2.3 Partially Deterministic Variables

In this section, we describe a new kind of distribution called partially deterministic. Partially deterministic distributions arise in the process of arc reversals in hybrid Bayesian networks.

The conditional distributions associated with a deterministic variable have values in units of probability mass. If some of the distributions have values in units of probability mass and some in units of probability density, we say that the distribution is *partially deterministic*. We get such distributions during the process of the arc

reversals between a continuous node and a deterministic node with discrete and continuous parents. Consider the Bayesian network shown on the left-hand side of Figure 2.19. Let ξ and ζ denote the continuous potentials at X and Z, respectively, before arc reversal, and ξ' and ζ' after arc reversal. Then,

$$\xi(x) = f(x),$$

$$\zeta(x, y, z, d_1) = [Z = X](x, z)\ (m),$$

$$\zeta(x, y, z, d_2) = [Z = Y](y, z)\ (m),$$

$$\zeta'(y, z, d_1) = (\xi \otimes \zeta)^{-X}(y, z, d_1) = \int f(x)\ [Z = X](x, z)\ dx = f(z),$$

$$\zeta'(y, z, d_2) = (\xi \otimes \zeta)^{-X}(y, z, d_2) = [Z = Y](y, z) \int f(x)\ dx = [Z = Y](y, z),$$

$$\xi'(x, y, z, d_1) = (\xi \otimes \zeta)\%(\xi \otimes \zeta)^{-X}(x, y, z, d_1) = [Z = X](x, z)\ ,$$

$$\xi'(x, y, z, d_2) = (\xi \otimes \zeta)\%(\xi \otimes \zeta)^{-X}(x, y, z, d_2) = f(x)\ [Z = Y](y, z)/[Z = Y](y, z)$$

$$= f(x)\ .$$

Thus, both X and Z have partially deterministic distributions.
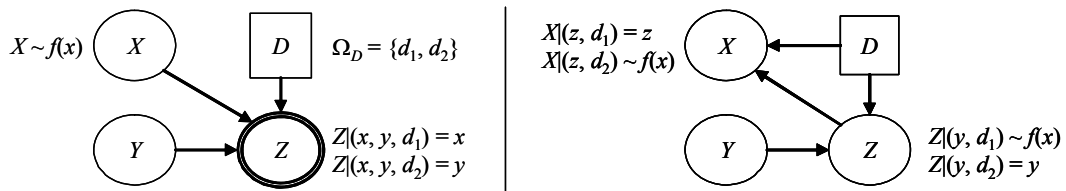
Figure 2.19. Arc reversal leading to partially deterministic distributions.

2.3   Converting non-MoG Bayesian Network to MoG Bayesian

Network

Consider the BN shown in Figure 1.2. It is not a MoG Bayesian network since D_5 has a non-Gaussian distribution, and C_{23} and F have non-linear conditional Gaussian distributions. This section explains how we can approximate a non-Gaussian distribution by a MoG distribution, and how we can approximate a max deterministic function by a MoG distribution.

2.3.1   Non-Gaussian Distributions

In this subsection, we will describe how the exponential distribution $E[1]$ can be approximated by a MoG distribution.

Let A denote a chance variable that has the exponential distribution with mean 1, denoted by $E[1]$, and let f_A denote its probability density function (PDF). Thus

$f_A(x) = e^{-x}$ if $0 \leq x$

$\quad = 0$ otherwise

In approximating the PDF f_A by a mixture of Gaussians, we first need to decide on the number of Gaussian components needed for an acceptable approximation. In this particular problem, more the components used, better will be the approximation. However, more components will lead to a bigger computational load in making inferences. We will measure the goodness of an approximation by estimating the Kullback-Leibler [1951] divergence measure between the target distribution and the corresponding MoG distribution.

Suppose we use five components. Then we will approximate f_A by the mixture PDF $g_A = p_1 \varphi_{\mu_1, \sigma_1} + \ldots + p_5 \varphi_{\mu_5, \sigma_5}$, where $\varphi_{\mu_i, \sigma_i}$ denote the PDF of a uni-variate Gaussian distribution with mean μ_i and standard deviation $\sigma_i > 0$, $p_1, \ldots, p_5 \geq 0$, and $p_1 + \ldots + p_5 = 1$. To estimate the mixture PDF, we need to estimate fourteen free parameters, e.g., $p_1, \ldots, p_4, \mu_1, \ldots, \mu_5, \sigma_1, \ldots, \sigma_5$. To find the values of the 14 free parameters, we solve a non-linear optimization problem as follows:

Find $p_1, \ldots, p_4, \mu_1, \ldots, \mu_5, \sigma_1, \ldots, \sigma_5$, so as to minimize $\delta(f_A, g_A)$

subject to: $p_1 \geq 0, \ldots, p_4 \geq 0, p_1 + \ldots + p_4 \leq 1, \sigma_1 \geq 0, \ldots, \sigma_5 \geq 0$,

where $\delta(f_A, g_A)$ denotes a distance measure between two PDFs. A commonly used distance measure is Kullback-Leibler divergence δ_{KL} defined as follows:

$$\delta_{KL}(f_A, g_A) = \int_S f_A(x) ln\left(\frac{f_A(x)}{g_A(x)}\right) dx$$

In practice, we solve a discrete version of the non-linear optimization problem by discretizing both f_A and g_A using a large number of bins. To discretize g_A, we assume that the domain of $\varphi_{\mu_i, \sigma_i}$ extends only from $\mu_i - 3\sigma_i$ to $\mu_i + 3\sigma_i$. With probability greater than 0.99, the domain of $E[1]$ extends from [0, 4.6]. To match the domain of the $E[1]$ distribution, we constrain the values $\mu_i - 3\sigma_i \geq 0$ and $\mu_i + 3\sigma_i \leq 4.6$ for $i = 1, \ldots, 5$. Suppose we divide the domain into n equally sized bins. Let f_i and g_i denote the probability masses for the ith bin corresponding to PDFs f_A and g_A, respectively. Then the discrete version of the non-linear programming problem can be stated as follows:

Minimize $\sum_{i=1}^{n} f_i(x) ln\left(\frac{f_i(x)}{g_i(x)}\right)$

subject to: $p_1 \geq 0, \ldots, p_4 \geq 0, p_1 + \ldots + p_4 \leq 1,$

$\sigma_1 \geq 0, \ldots, \sigma_5 \geq 0,$

$\mu_1 - 3\sigma_1 \geq 0, \ldots, \mu_5 - 3\sigma_5 \geq 0,$

$\mu_1 + 3\sigma_1 \leq 4.6, \ldots, \mu_5 + 3\sigma_5 \leq 4.6$

One can use the solver in Excel to solve such optimization problems taking care to avoid local optimal solutions. An optimal solution computed in Excel with $n = 100$ (shown rounded to 3 digits) is shown in Table 2.3.

Table 2.3: Parameters of the MoG Approximation to the $E[1]$ distribution.

i	p_i	μ_i	σ_i
1	0.051	0.032	0.011
2	0.135	0.143	0.048
3	0.261	0.415	0.138
4	0.341	1.014	0.338
5	0.212	2.300	0.767

A graph of the two PDFs overlaid over each other is shown in Figure 2.20.
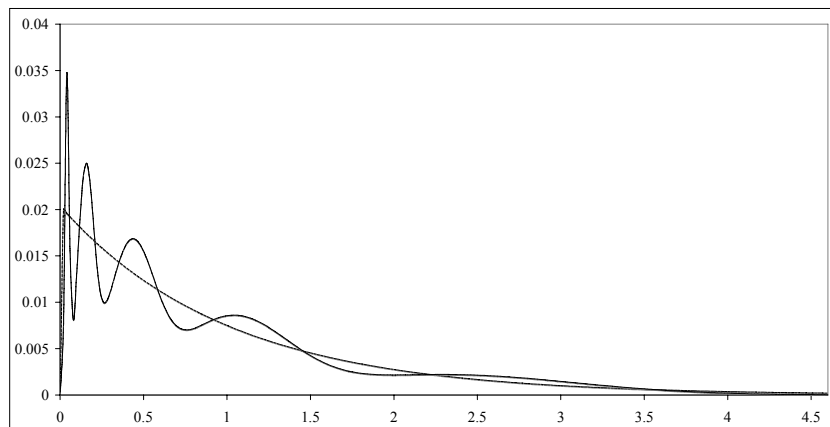
Figure 2.20. A 5-component MoG approximation (solid) of the *E*[1] distribution

(dashed).

To measure the goodness of the approximation, we can compute the Kullback-Leibler (KL) divergence of the two distributions over the domain [0, 4.6] where both densities are positive. The KL divergence is approximately 0.021. We can also compare moments. The mean and variance of the *E*[1] distribution are 1 and 1. The mean and variance of the MoG approximation are 0.96 and 0.76. We can also compare the cumulative distribution functions (CDF). A graph of the two CDFs overlaid over each other is shown in Figure 2.21.
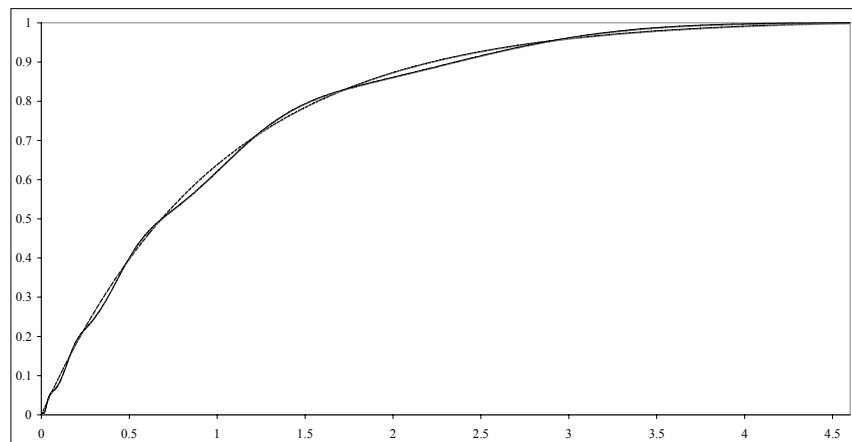
Figure 2.21: The CDFs of the $E[1]$ distribution (dashed), and its MoG approximation

(solid).

If we need to get a MoG approximation of the $E[\lambda]$ distribution, we can derive it easily from the MoG approximation of the $E[1]$ distribution. If $X \sim E[1]$, and $Y = \lambda X$, then $Y \sim E[\lambda]$. Thus, to get a MoG approximation of $E[25]$, e.g., the mixture weights p_i's don't change, but we need to multiply each mean μ_i and σ_i in Table 2.3 by 25.

2.3.2   Maximum of Two Gaussians

The problem of computing the distribution of the maximum of two (or more) Gaussians is of interest in many communities, especially in project management [Clark 1961], statistics [Afonja 1972], and in design of semiconductors [Sinha *et al*. 2006]. Clark [1961] computes exactly the first four moments of the maximum of two correlated Gaussians. To compute the moments of maximum of three or more correlated Gaussians, he makes the assumption that the distribution of the maximum is Gaussian, which allows a recursive computation. This assumption is grossly

violated, especially if the difference of the means is much smaller than the variances of the two Gaussians.

Here, we will represent the maximum of two Gaussians as a Bayesian network and approximate it with a mixture of Gaussians (MoG) Bayesian network. Lauritzen and Jensen's [2001] exact algorithm for computing marginals in MoG Bayesian network can then be used to compute the full distribution of the maximum.

Consider the Bayesian network shown in Figure 2.22 with three random variables: $E \sim N(5, 1/16)$, $F \sim N(5, 1)$, E and F are independent, and $G = Max\{E, F\}$. Since the deterministic function is not linear, this Bayesian network is not a MoG.
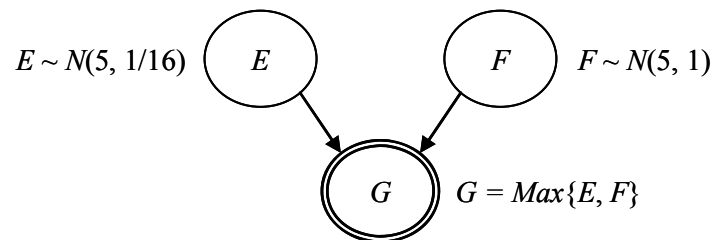
Figure 2.22: Maximum of two Gaussians Bayesian network

For this small BN, we can compute the marginal probability density function of G by brute force using order statistics. Let F_G denote the cumulative distribution function (CDF) of G, Φ_F denote the CDF of F, and let Φ_E denote the CDF of E. Then, $F_G(g) = \Phi_E(g)\Phi_F(g)$. Therefore, the probability density function of G, f_G, is given by $f_G(g) = (d/dg)(F_G(g)) = \phi_E(g)\Phi_F(g) + \Phi_E(g)\phi_F(g)$, where ϕ_E and ϕ_F are the PDFs of E and F, respectively. Since there is no closed form expression for the CDF of a normal distribution, there is no closed form expression for $f_G(g)$. A graph of $f_G(g)$ is shown in

Figure 2.23. The mean and variance of G are computed (using the exact analytic results given by Clark [1961]) as 5.411 and 0.362, respectively.
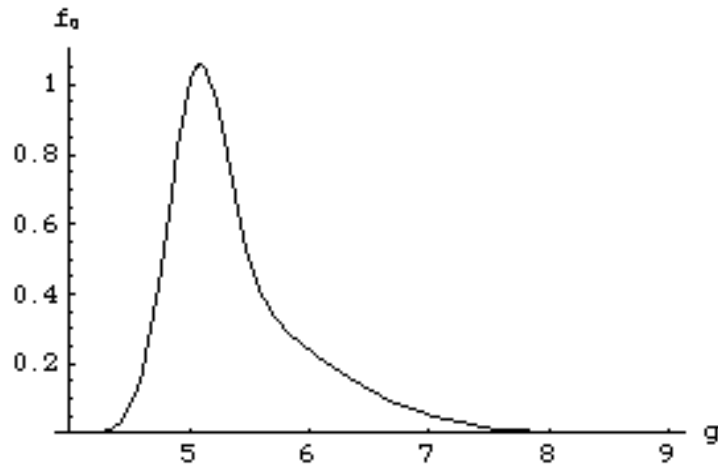
Figure 2.23: The probability density function of G.

Our main strategy is to compute the marginal distribution of G using local computation. For that we will convert the BN in Figure 2.22 to a mixture of Gaussians (MoG) BN. Then we can use the Lauritzen-Jensen algorithm [2001] for MoG BN to compute the marginal distribution of G using local computation. This algorithm is implemented in Hugin, a commercially available software package.

In approximating a Bayesian Network with MoG BN, we will assume that the *effective* domain of a univariate Gaussian distribution with mean μ and standard deviation σ is $(\mu - 3\sigma, \mu + 3\sigma)$. Thus, the effective domain of E is (4.25, 5.75) and the effective domain of F is (2, 8). Since $G = Max\{E, F\}$, the effective domain of G is (4.25, 8).

Our first step is to introduce a new discrete random variable S as shown in Figure 2.24. S has two states s_1 and s_2, and has E and F as parents. The conditional

distributions of S are as follows: $P(s_1|e, f) = [E \le F](e, f)$ and $P(s_2|e, f) = [E > F](e, f)$. Thus, S is an indicator random variable which is in state s_1 when $E \le F$, and in state s_2 when $E > F$. We make S a parent of G, and the conditional distributions of G can now be expressed as a conditional linear Gaussian as shown in Figure 2.24. Notice that the conditional distributions of all continuous variables are conditional linear Gaussians. The Bayes net is not a MoG Bayes net since S is a discrete random variable with continuous parents E and F.
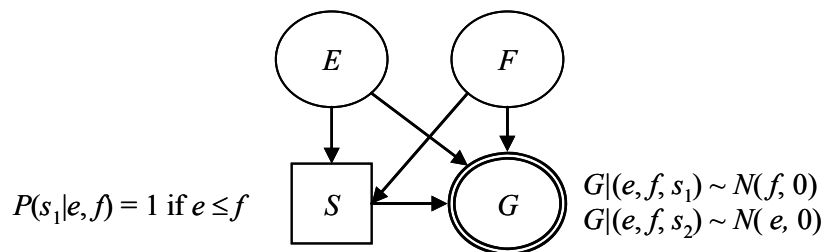
Figure 2.24: The augmented Bayesian network with discrete random variable S.

Our next step is to do a sequence of arc reversals so that the resulting Bayes net is an equivalent MoG Bayes net. We need to reverse arcs (E, S) and (F, S) in either sequence. Suppose we reverse (E, S) first.

Before arc reversal, let ε denote the density potential at E, and let σ denote the discrete potential at S. Thus, $\varepsilon(e) = \varphi_{5,1/4}(e)$, where $\varphi_{5,1/4}(e)$ denotes the probability density function (PDF) of a univariate Gaussian distribution with mean 5, and standard deviation ¼, and $\sigma(e, f, s_1) = [E \le F](e, f)$, $\sigma(e, f, s_2) = [E > F](e, f)$

After arc reversal the revised potential at S is $\sigma' = (\varepsilon \otimes \sigma)^{-E}$, and the revised potential at E is $\varepsilon' = (\varepsilon \otimes \sigma)\%(\varepsilon \otimes \sigma)^{-E}$. The details of these potentials are as follows.

$$\sigma'(f, s_1) = P(s_1|f), \text{ where } P(s_1|f) = \int_{-\infty}^{f} \varphi_{5,1/4}(e)de, \sigma'(f, s_2) = P(s_2|f), \text{ where } P(s_2|f)$$

$$= \int_{f}^{\infty} \varphi_{5,1/4}(e)de, \varepsilon'(e,f, s_1) = (1/P(s_1|f)) \varphi_{5,1/4}(e)[E \le F](e,f), \varepsilon'(e,f, s_2) = (1/P(s_2|f))$$

$$\varphi_{5,1/4}(e)[E > F](e,f)$$
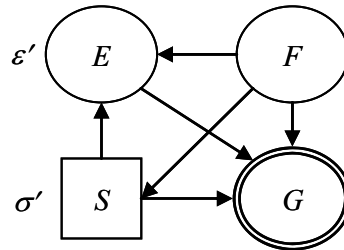
Figure 2.25: The Bayes net after reversal of arc (E, S).

A graph of $P(s_1|f)$ vs. f is shown in Figure 2.26. Notice that since $P(s_1|f)$ is the cumulative distribution function of E, $P(s_1|f) \approx 0$ if $f < 4.25$, ≈ 1 if $f > 5.75$. Figure 2.27 shows the conditional PDF of E given s_1 and some sample values of f. Notice that these are truncated Gaussians.

Notice that the conditional probability densities of E (given F and S) are no longer conditional linear Gaussians. Later (after we have reversed arc (F, S)), we will approximate these conditional distributions by mixtures of Gaussians.
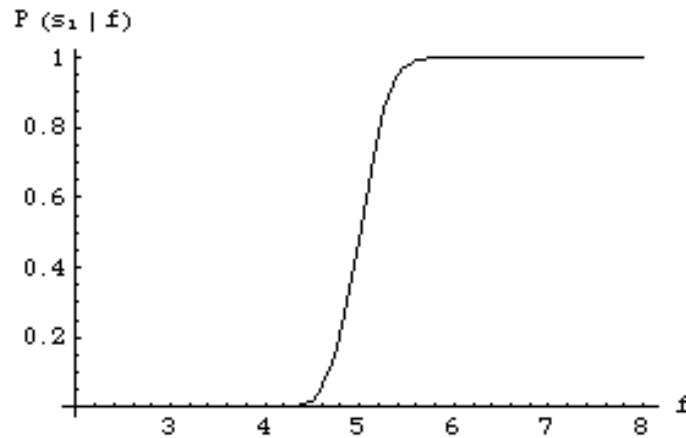
Figure 2.26: A graph of $P(s_1 \mid f)$ versus f
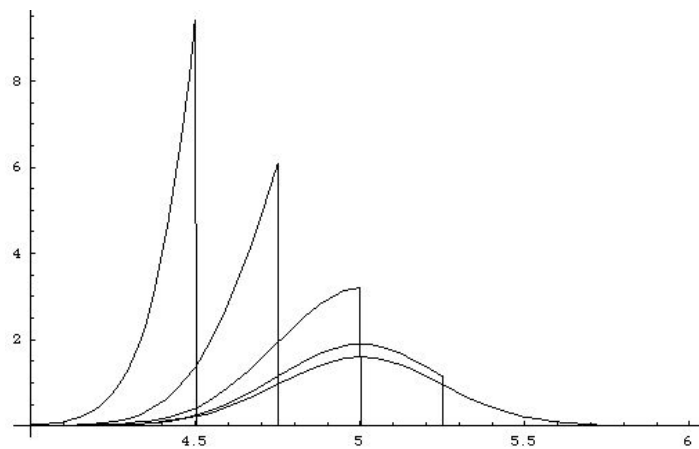
Figure 2.27: A graph of $f_{E\mid(s_1,f)}(e)$ vs. e for $f = 4.5$, 4.75, 5, 5.25, and 6

Next, we reverse arc (F, S). Suppose the density potential at F before arc reversal is denoted by ϕ. Then the revised potential at S is $\sigma''(s_1) = 0.5$, $\sigma''(s_2) = 0.5$, and the revised potential at F is $\phi'(f, s_1) = 2P(s_1|f)\varphi_{5,1}(f))$, $\phi'(f, s_2) = 2P(s_2|f)\varphi_{5,1}(f)$. The potential ϕ' represents conditional probability densities for F given s_1 and s_2. A graph of these two densities is shown in Figure 2.28. Clearly, these are not conditional linear Gaussians. The revised Bayes net is shown in Figure 2.29.
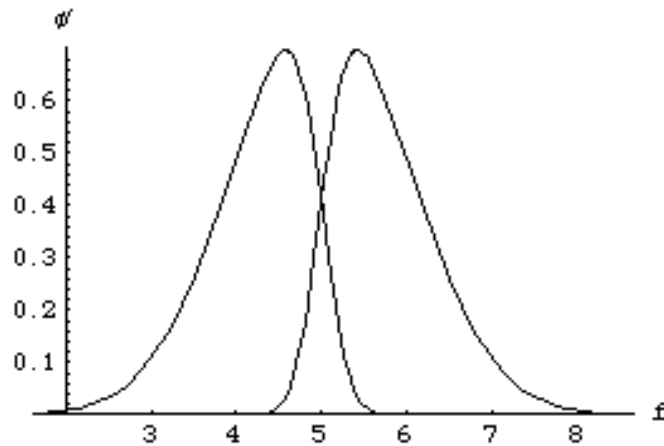
Figure 2.28: The conditional probability densities of F given s_1 (right) and s_2 (left)
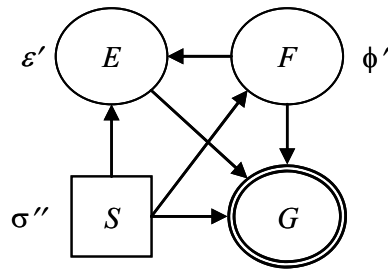
Figure 2.29: The Bayes net after reversal of arc (F, S)

Notice that the Bayes net in Figure 2.29 is almost a MoG BN except for the fact that the potentials ϕ' and ε' are not conditional linear Gaussians. We can approximate these potentials by mixtures of Gaussian potentials using the optimization technique described in Shenoy [2006].

When the non-Gaussian distribution has many continuous variables in its domain, the task of finding a mixture of Gaussians approximation can be difficult in practice. Notice from Figure 2.27 that the potential for ε' changes considerably with f. So our next step is to introduce a new discrete variable D with $\{F, S\}$ as its parents and E as its child. D has seven states, $d_1, \ldots, d_7$. D is an indicator variable for $F|s_i$. Given s_1, d_1

$\equiv -\infty < f \leq 4.5,\ d_2 \equiv 4.5 < f \leq 4.75,\ \dots,\ d_6 \equiv 5.5 < f \leq 5.75,$

$d_7 \equiv 5.75 < f < \infty.$ Given $s_2,\ d_1 \equiv 5.5 < f < \infty,\ d_2 \equiv 5.25 < f \leq 5.5,\ \dots,$

$d_6 \equiv 4.25 < f \leq 4.5,\ d_7 \equiv -\infty < f < 4.25.$ The new Bayes Net is shown in Figure 2.23.

The potential ε' at E changes its domain from $\{E, F, S\}$ to $\{E, F, S, D\}$, but otherwise

remains unchanged.
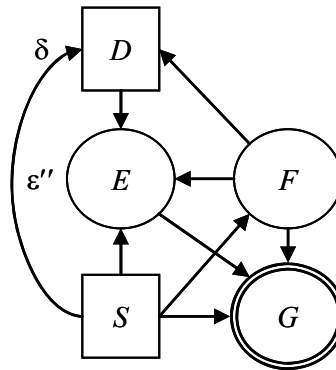
Figure 2.30: The Bayes net after adding the discrete variable D

Notice that D is a discrete variable with continuous parent F. We need to address

this situation by reversing the arc (F, D). The revised discrete potential at D is

$\delta'(d_1,\ s_1) = \delta'(d_1,\ s_2) = \int_{-\infty}^{4.5} 2P(s_1 \mid f)\, \varphi_{5,1}(f)df = 0.0014,$ etc. The revised density

potential at F is as follows.

$\phi''(f, s_1, d_1) = (2P(s_1 | f)/0.0014)\, \varphi_{5,1}(f))[-\infty < F \leq 4.5](f),$

$\phi''(f, s_2, d_1) = (2P(s_2 | f)/0.0014)\, \varphi_{5,1}(f))\, [5.5 < f < \infty],$ etc.

The resulting Bayes net after reversing the arc (F, D) is shown in Figure 2.31.
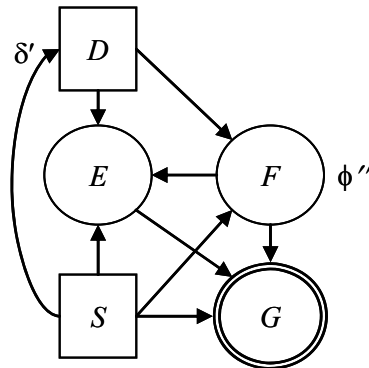
Figure 2.31: The Bayes net after reversing arc (F, D)

Next we need to approximate the non-CLG potentials ϕ'' and ε'' by CLG potentials as described in Shenoy [2006]. In order to approximate ϕ'', we use five Gaussian components for each region d_i. Figure 2.32 shows the actual distribution of $F|(s_1, d_1)$ overlaid on a MoG approximation, and Figure 2.33 shows the revised Bayes net. The potential χ at discrete variable C has the weights of the MoG approximation, and the potential ϕ''' has the parameters of the Gaussian components.
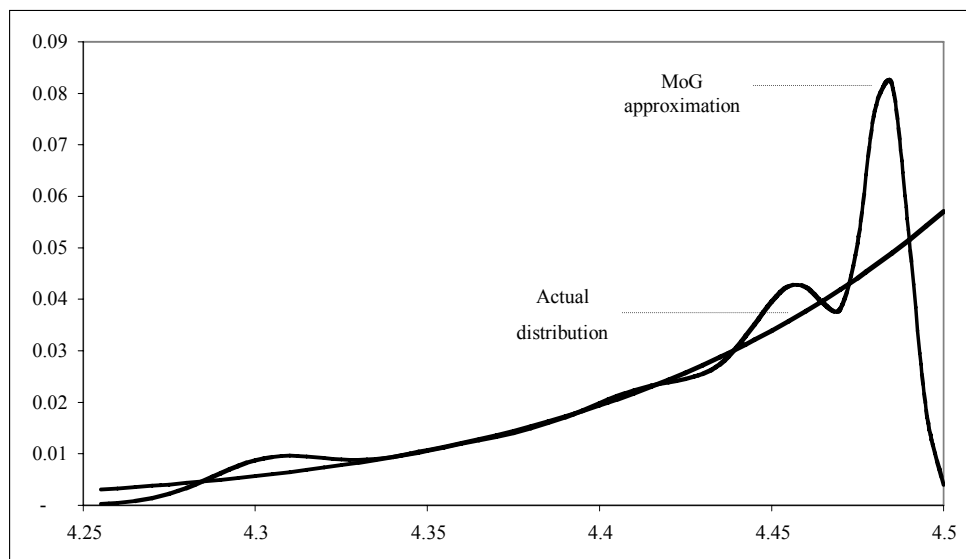
Figure 2.32: A MoG approximation of pdf of $F|(s_1, d_1)$
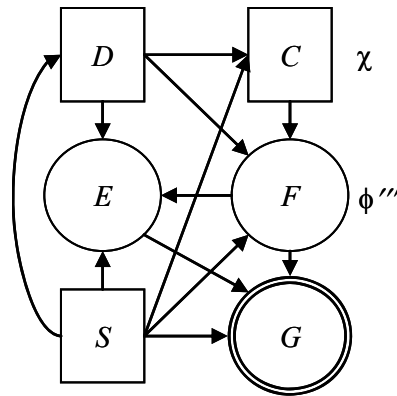
Figure 2.33: The Bayes net after approximating the potential ϕ'' with MoG

distributions

Finally, we approximate the potential ε'' by a MoG distribution. The task of approximating the conditional distribution of $E|(s_i, f, d_j)$ is much simpler since range in which F lies is restricted to a small region. For example $E|(s_1, f, d_7) \sim N(5, 1/16)$, and no approximation is needed. For the remaining regions, we use 3 components. Figure 2.34 shows a MoG approximation using three Gaussian components for $f = 4.25, 4.30, \ldots, 4.5$ for $E|(s_1, f, d_1)$ overlaid on the actual distribution. The Gaussian components are correlated with F. Figure 2.35 shows the final MoG Bayes net. The potential associated with A has the weights of the mixtures, and the potential ε''' has the parameters of the Gaussian components including the correlations with F.
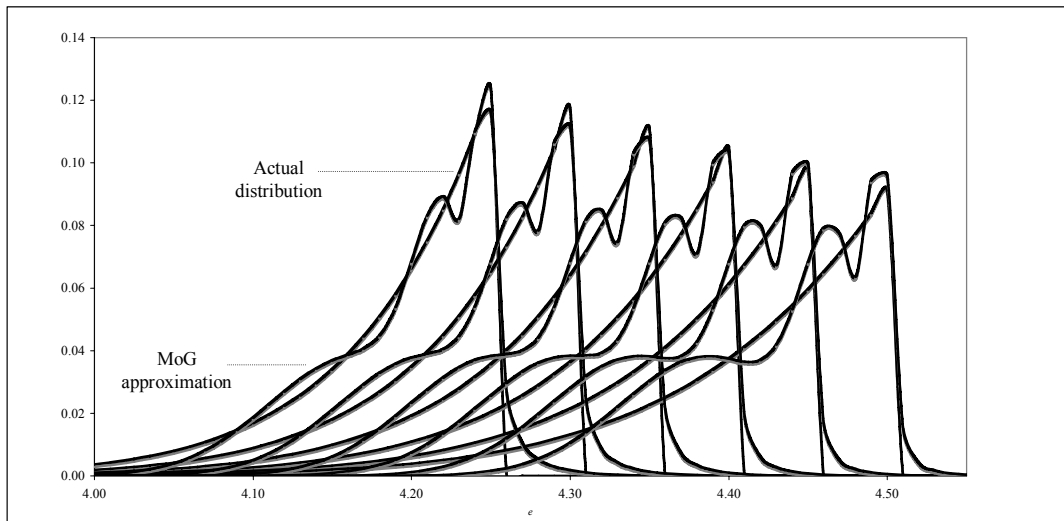
Figure 2.34: A MoG approximation of conditional distribution of E given f, s_1, and d_1
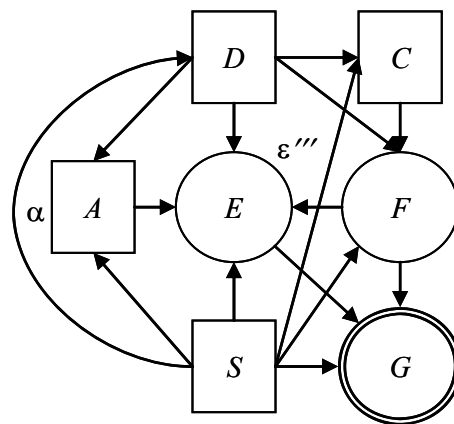
Figure 2.35: The MoG Bayes net after approximating ε'' with MoG distributions

To evaluate the quality of our approximation we entered the above MoG BN in Hugin. The mean of the marginal distribution of G is reported as 5.409 and the variance is reported as 0.351. Comparing to 5.411 and 0.362 given by the exact analytic results, the approximation is very good considering that we used a few components for approximating the potentials at F and E. The full distribution of G computed in Hugin is shown in Figure 2.36 below.
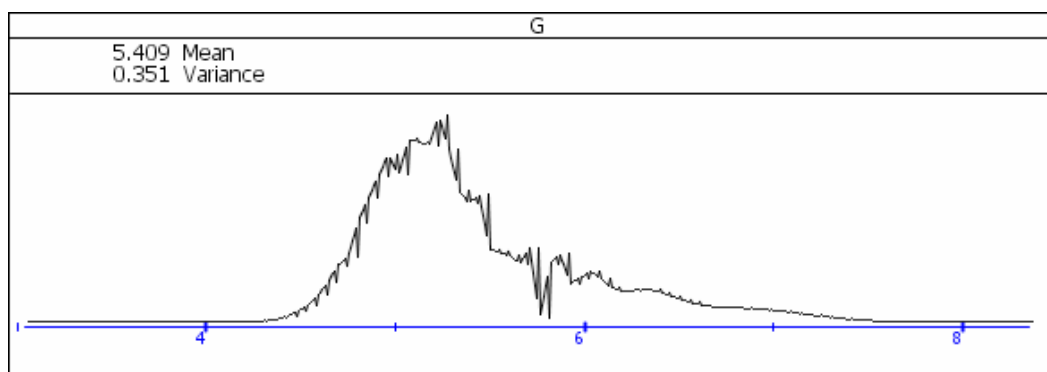
Figure 2.36: The full distribution of *G* in Hugin

2.4   Summary and Conclusions

In chapter one the problems involved with the current methods of solving stochastic PERT networks were illustrated. In this chapter, we provide a new method which aims to approximate the true distribution of the project completion time by eliminating the false assumptions for the distribution of the maximum of two Gaussians. In this method first a PERT network is transformed into a MoG Bayesian network and then the Lauritzen-Jensen algorithm is used to make exact inferences in the MoG Bayes net.

To transform a PERT network into a MoG Bayesian network the non-Gaussian distributions and max deterministic functions should be approximated using MoG distributions. These two cases are illustrated in subsections 2.3.1 and 2.3.2 respectively. In addition, transformation of a PERT Bayes net into a MoG involves arc reversals since the restrictive nature of MoG distributions does not allow discrete variables to have continuous parents which is in fact is very common for general

hybrid Bayesian networks. Also in the process of approximating non-CLG distributions with MoG's we may create discrete variables with continuous parents. Besides, arc reversals are also used to solve hybrid influence diagrams, so there is a need for arc reversal theory in literature. For that reason, in Section 2.2 we have described arc reversals in hybrid Bayesian networks with deterministic variables for all possible cases. We also have described a new kind of variable called partially deterministic that can arise after arc reversals. The arc reversal theory facilitates the task of approximating general Bayesian networks with mixtures of Gaussians Bayesian networks.

Some disadvantages of our strategy are as follows. Arc reversals make Gaussian distributions non-Gaussian. We can approximate non-Gaussian distributions by mixtures of Gaussians. However, when the non-Gaussian distribution has many continuous variables in its domain, the task of finding a mixture of Gaussians approximation by solving an optimization problem can be difficult in practice. Additionally, in the process of arc reversals, we increase the domains of the potentials and the resulting complexity may make this strategy impractical. Transformation process of the PERT Bayes net given in Figure 1.2 into a MoG Bayesian network is demonstrated in Appendix A.

In Chapter 3 we will explore solving stochastic PERT networks using Mixtures of Truncated Exponentials as an alternative to transforming PERT networks into a MoG Bayesian network.

3 SOLVING STOCHASTIC PERT NETWORKS USING MIXTURES OF TRUNCATED EXPONENTIALS

3.1 Introduction

In the previous chapters we transformed a PERT network into a Bayesian network, which enabled us to model the dependence between the activities. Following that we demonstrated the methods to transform a PERT Bayesian network into a MoG Bayesian network. Representation of a PERT network as a MoG Bayesian network is beneficial in the sense that it allows us to eliminate the false assumption made in the literature which assumes that the maximum of two normally distributed independent random variables is again normally distributed. Hence, we concluded that by transforming PERT networks into MoG Bayesian network a more accurate representation of the project completion time is feasible. However, as discussed in the previous chapter the transforming process of a PERT network into a MoG Bayesian network is cumbersome because of the restricted nature of MoG Bayesian networks. The inability of discrete variables to have continuous parents and the enforcement for continuous variables to possess conditional linear Gaussian distributions makes the process of transforming a PERT network to a MoG Bayes net too complex for practical use.

An alternative to solve stochastic PERT networks with mixtures of Gaussians is to solve them using mixtures of truncated exponentials. Mixtures of truncated exponentials (MTE) are an alternative to discretization and Monte Carlo methods for

solving hybrid Bayesian networks. MTE potentials can be used for inference in hybrid Bayesian networks that do not fit the restrictive assumptions of the conditional linear Gaussian (CLG) model, such as networks containing discrete nodes with continuous parents. The following section provides detailed information about mixtures of truncated exponentials.

3.2   Mixtures of Truncated Exponentials

A mixture of truncated exponential (MTE) [Moral *et al.* , 2001; Rumi, 2003] has the following definition.

Let X be a mixed n-dimensional random variable. Let $Y = (Y_1, \ldots, Y_d)$ and $Z = (Z_1, \ldots, Z_c)$ be the discrete and continuous parts of X, respectively, with $c + d = n$. A function $\phi \colon \Omega_X \mapsto R^+$ is an MTE potential if one of the next two conditions holds: The potential ϕ can be written as

$$\phi(x) = \phi(y, z) = a_0^y + \sum_{i=1}^{m} a_i^y \exp\left(\sum_{j=1}^{c} b_j^y z_j\right) \qquad (3.1)$$

where a_0^y, a_i^y and b_j^y are real numbers for all $i = 1, \ldots, m$, $j = 1, \ldots, c$, $y \in \Omega_Y$ and $z \in \Omega_Z$.

There is a partition $\Omega_1, \ldots, \Omega_k$ of Ω_X verifying that the domain of continuous variables, Ω_Z, is divided into hypercubes, the domain of the discrete variables, Ω_Y, is divided into arbitrary sets, and such that ϕ is defined as

$$\phi(x) = \phi_i(x) \qquad \text{if } x \in \Omega_i,$$

where each ϕ_i, $i = 1, \ldots, k$ can be written in the form of equation (3.1)

In the definition above, k is the number of pieces and m is the number of exponential terms in each piece of the MTE potential.

Mixtures of truncated exponentials (MTE) are an alternative to discretization and Monte Carlo methods for solving hybrid Bayesian networks. Any probability density function can be approximated by an MTE potential, which can always be marginalized in closed form.

Consider a normally distributed random variable X with mean μ and variance $\sigma^2 > 0$. The PDF for the normal distribution is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-1/2\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

A general formulation for a 2-piece, 3-term unnormalized MTE potential which approximates the normal PDF is as follows [Cobb and Shenoy, 2006a].

$$\psi'(x) = \begin{cases} \sigma^{-1}(-0.010564 + 197.055720\exp\{2.2568434(\frac{x-\mu}{\sigma})\} \\ -461.439251\exp\{2.3434117(\frac{x-\mu}{\sigma})\} \\ +264.793037\exp\{2.4043270(\frac{x-\mu}{\sigma})\}) \qquad if \quad \mu - 3\sigma \le x < \mu \\ \sigma^{-1}(-0.010564 + 197.055720\exp\{-2.2568434(\frac{x-\mu}{\sigma})\} \\ -461.439251\exp\{-2.3434117(\frac{x-\mu}{\sigma})\} \\ +264.793037\exp\{-2.4043270(\frac{x-\mu}{\sigma})\}) \qquad if \quad \mu - 3\sigma \le x < \mu \\ 0 \quad otherwise \end{cases} \qquad (3.2)$$

The properties of the MTE potential in (3.2) are as follows:

(1) $\displaystyle\int_{\mu-3\sigma}^{\mu+3\sigma} \psi'(x)\,dx = \int_{\mu-3\sigma}^{\mu+3\sigma} f_X(x)dx = 0.9973$

(2) $\displaystyle\int_{\mu-3\sigma}^{\mu+3\sigma} \psi'(x)dx = \int_{\mu}^{\mu+3\sigma} \psi'(x)dx = \int_{\mu-3\sigma}^{\mu} f_X(x)dx = \int_{\mu}^{\mu+3\sigma} f_X(x)dx = 0.49865$

(3) $\psi' \geq 0$

(4) $\psi'(x)$ is symmetric around μ.

A normalized version of the 2-piece, 3-term MTE approximation to the normal PDF is as follows:

$$\psi(x) = (1/0.9973)*\psi'(x). \qquad (3.3)$$

Following are the properties of the normalized MTE potential in (3.3):

(1) $\displaystyle\int_{\mu-3\sigma}^{\mu+3\sigma} \psi(x)dx = 1$

(2) $\displaystyle\int_{\mu-3\sigma}^{\mu+3\sigma} x\psi(x)dx = \mu$

(3) $\displaystyle\int_{\mu-3\sigma}^{\mu+3\sigma} (x-\mu)^2\psi(x)dx = 0.98187\sigma^2$

MTE potentials are a nice alternative to Mixtures of Gaussians as an inference tool in hybrid Bayesian networks. MTE potentials can be used for inference in hybrid Bayesian networks that do not fit the restrictive assumptions of the conditional linear Gaussian (CLG) model. Mixtures of truncated exponentials Bayesian networks with linear deterministic variables can be solved easily using the Shenoy-Shafer architecture for finding marginals of large multivariate distributions represented as a

Bayesian network.[Cobb and Shenoy, 2005]. PERT Bayesian networks may have max deterministic nodes which are nonlinear. But we will show that such networks can also be solved using MTE's. In the following section we will transform a PERT network into a PERT Bayesian network which we later going to approximate using mixtures of truncated of exponentials. As the last step we are going to solve the resulting MTE PERT Bayesian network using the Shenoy-Shafer architecture.

3.3 Solving a PERT Network Using Mixtures of Truncated Exponentials

3.3.1 Representation of a PERT network as a Bayesian network

Consider the PERT network given in Figure 3.1. This network represents a project with the activities A_1, A_2 and A_3. S stands for the project start time and E stands for the project completion time. We assume that the project start time is zero. The precedence constraints, represented by arcs, are as follows: The activities A_1 and A_2 do not have any predecessors. The activity A_3 can only be started after A_1 is completed. The project is completed after all three activities are completed.
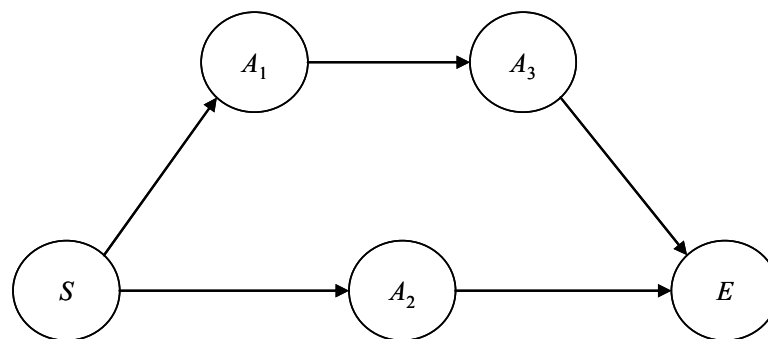
Figure 3.1. An Example of a stochastic PERT network with three activities

The distributions of activity durations are known, and we are informed that the activity durations A_1 and A_3 are positively correlated. Following the method described in Jenzarli[1995] this PERT network will be transformed into a PERT Bayesian network which allows us to depict the dependencies between the activity durations.

Let D_i and C_i denote the duration of the activity i. As the first step of the transformation process, the activity durations will be replaced with activity completion times. The resulting network is given in Figure 3.2.
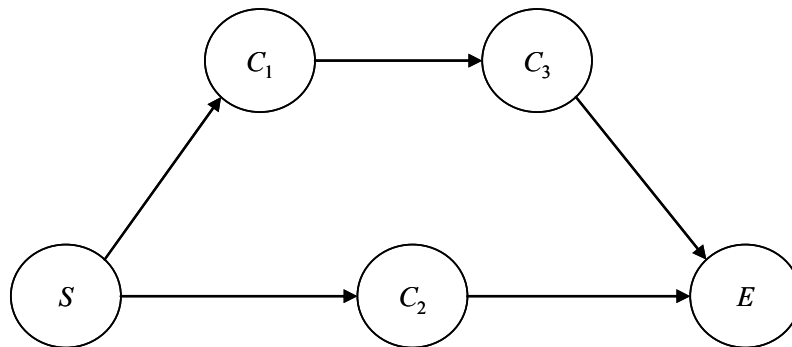
Figure 3.2. The PERT example after the activities are replaced with activity

completion times

As the next step, activity durations will be added with an with an arrow from D_i to C_i, so that each activity will be represented by two nodes, its duration D_i and its completion time C_i. The resulting network is depicted in Figure 3.3.
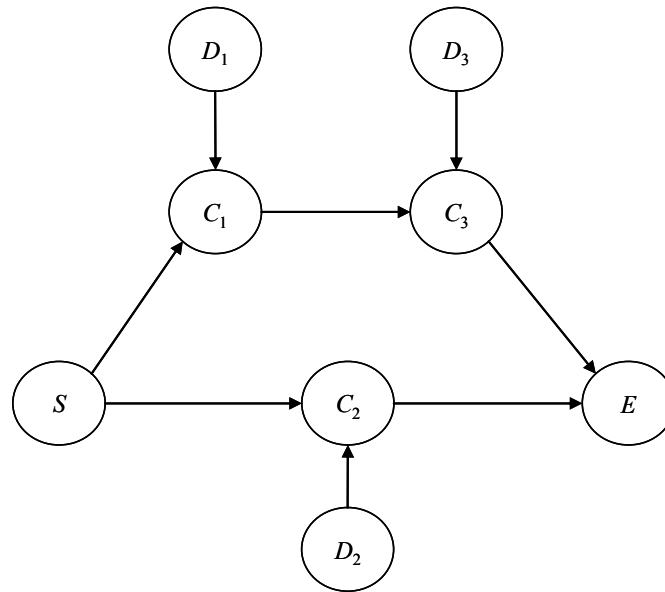
Figure 3.3. The PERT example after each activity is represented through their

duration and completion times

Notice that the completion time of the activities which do not have any predecessors will be the same as their durations. Hence, as the next step of the transformation process, the activities A_1 and A_2 are going to be represented just by their durations, as D_1 and D_2. We assume that the project start time is zero with probability 1. The resulting network is given in Figure 3.4 below. We assume that each activity will be started as soon as all the preceding activities are completed. Accordingly, E represents the completion time of the project, which is the $Max\{D_2, C_3\}$.
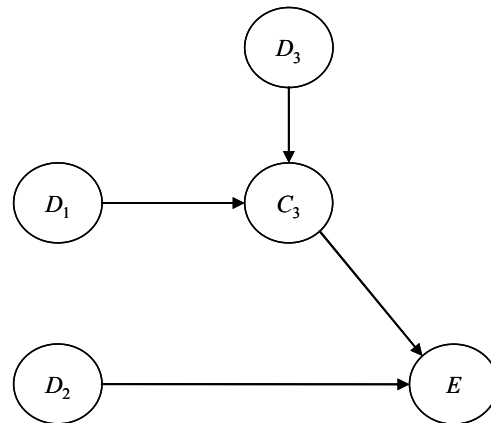
Figure 3.4. The PERT example after the activities without any predecessors are represented just with their durations.

As the last step of the transformation process the dependency between the activities will be depicted. In this example we are informed that the duration of the activities D_1 and D_3 are positively correlated. In order to depict this dependency we add an arc from D_1 to D_3. The resulting PERT Bayes net is given in Figure 3.5 below. Notice that the deterministic variables, C_3 and E, are depicted as double bordered ovals.

$$C_3 = D_1 + D_3$$

$$E = Max\{D_2, C_3\}$$
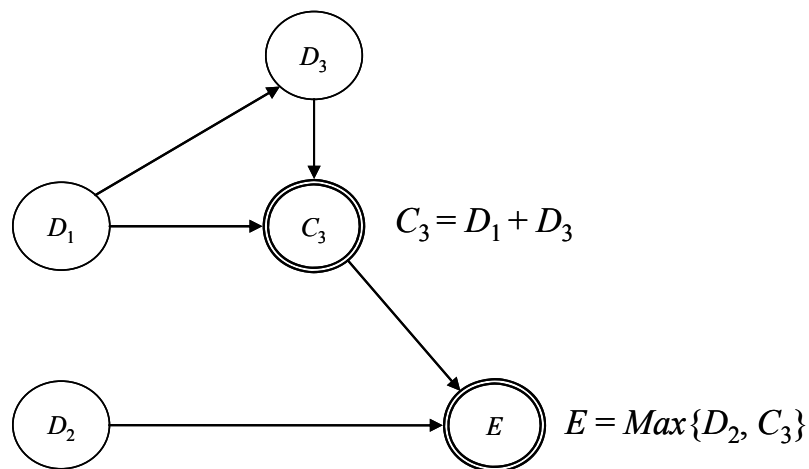
Figure 3.5. The resulting PERT Bayesian network

3.3.2 Approximation of Activity Distributions Using Mixtures of Truncated Exponentials

The primary objective of this study is to compute the completion time of the project without setting any assumptions for activity distributions. This objective will be materialized by approximating the activity durations using mixtures of truncated exponentials and propagating the resulting mixtures of truncated exponentials network using the Shenoy-Shafer architecture.

Consider the PERT Bayes Net given in Figure 3.6. Notice that it is not a mixtures of exponentials Bayesian network since the activity durations D_1, D_2, and D_3 are all normally distributed. In order to transform the PERT Bayes net into a MTE Bayesian network those activities need to be approximated using MTE's.

$$D_3 \sim N(0.6 + D_1, 0.04)$$

$$D_1 \sim N(0.4, 0.01)$$

$$C_3 = D_1 + D_3$$

$$D_2 \sim N(1.4, 0.01)$$

$$E = \text{Max}\{D_2, C_3\}$$
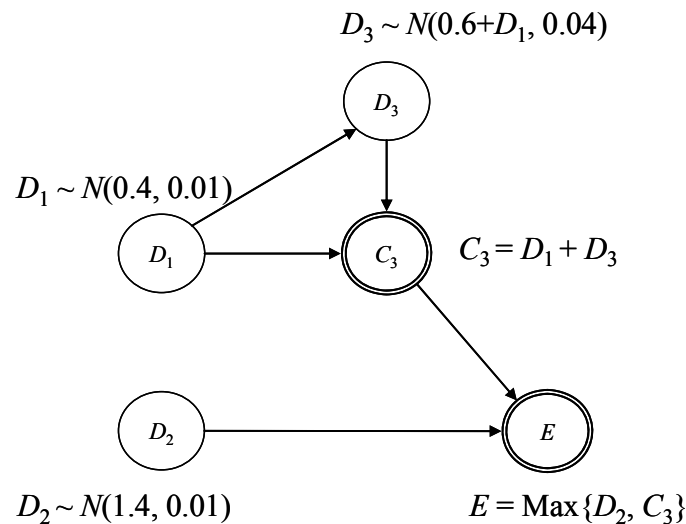
Figure 3.6: An example of a PERT Bayesian network

The probability distribution for D_1 is defined as $D_1 \sim N(0.4, 0.01)$. The PDF for D_1 is approximated by a 2-piece 3 term MTE potential as follows:

$$\delta_1(d_1) = \begin{cases} 10.027(-0.010564+197.056\exp\{22.5684(-0.4+d_1)\} \\ -461.439\exp\{23.4341(-0.4+d_1)\}+264.793\exp\{24.0433(-0.4+d_1)\})\} \\ \textit{if } 0.1 \le d_1 < 0.4 \\ \\ 10.027(-0.010564+264.793\exp\{-24.0433(-0.4+d_1)\} \\ -461.439\exp\{-23.4341(-0.4+d_1)\}+197.056\exp\{-22.5684(-0.4+d_1)\})\} \\ \textit{if } 0.4 \le d_1 \le 0.7 \end{cases}$$

The MTE approximation of D_1 overlaid on the actual normal distribution is depicted in Figure 3.7 below.
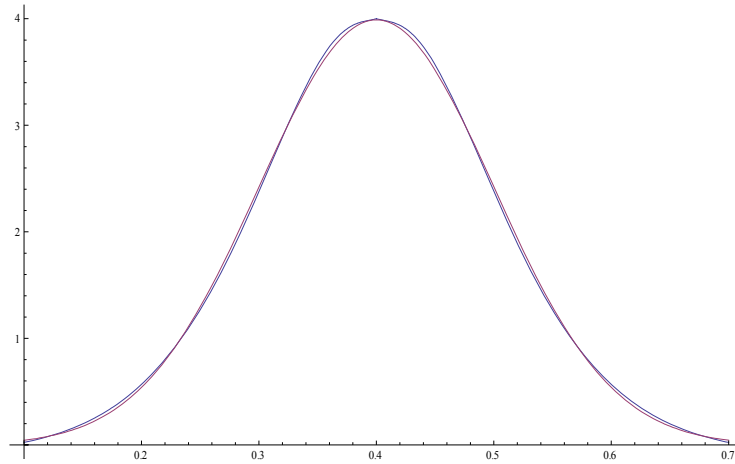
Figure 3.7. The actual distribution of D_1 overlaid on its MTE approximation

The probability distribution for D_2 is defined as $D_2 \sim N(1.4, 0.01)$. The PDF for D_1 is approximated by an MTE potential as follows:

$$\delta_2(d_2) = \begin{cases} 10.027(-0.010564+197.056\exp\{22.5684(-1.4+d_2)-461.439 \\ \exp\{23.4341(-1.4+d_2)+264.793\exp\{24.0433(-1.4+d_2)\})\} \\ \textit{if } 1.1 \le d_2 < 1.4 \\ \\ 10.027(-0.010564+264.793\exp\{-24.0433(-1.4+d_2)-461.439 \\ \exp\{-23.4341(-1.4+d_2)+197.056\exp\{-22.5684(-1.4+d_2)\})\} \\ \textit{if } 1.4 \le d_2 \le 1.7 \end{cases}$$

The MTE approximation of D_2 overlaid on the actual distribution is depicted in Figure 3.8 below.
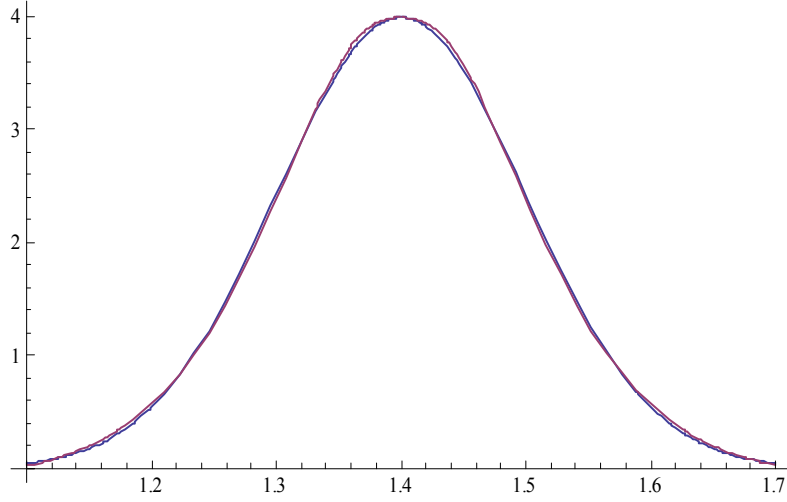
Figure 3.8. The actual distribution of D_2 overlaid on its MTE approximation

The probability distribution for D_3 is defined as $D_3|d_1 \sim N(0.6+d_1, 0.04)$. The conditional PDF for $D_3|d_1$ is approximated by an MTE potential as follows:

$$
\delta_3(d_3,d_1)=\begin{cases}
5.0135(-0.010564+197.056\exp\{11.2842(-0.6-d_1+d_3)\}-461.439 \\
\exp\{11.7171(-0.6-d_1+d_3)+264.793\exp\{12.0216(-0.6-d_1+d_3)\}\}) \\
if \quad (d_1 \le d_3 < 0.6+d_1)\,\&\,(0.1 \le d_1 < 0.7) \\
\\
5.0135(-0.010564+264.793\exp\{-12.0216(-0.6-d_1+d_3)\}-461.439 \\
\exp\{-11.7171(-0.6-d_1+d_3)+197.056\exp\{-11.2842(-0.6-d_1+d_3)\}\}) \\
if \quad (0.6+d_1 \le d_3 \le 1.2+d_1)\,\&\,(0.1 \le d_1 < 0.7)
\end{cases}
$$

The plot for the MTE approximation for D_3 is given in Figure 3.9 below.
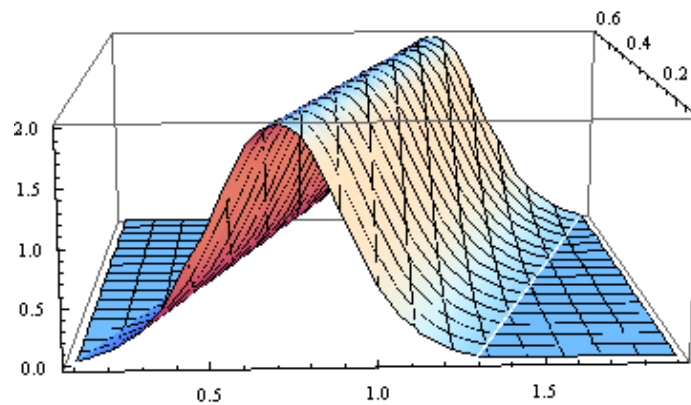
Figure 3.9. MTE approximation for $D_3|d_1$

3.3.3 Operations in MTE Networks

In section 3.3.1 we transformed an example of a PERT network into a PERT Bayesian network. Following that, in section 3.3.2, we approximated the activity distributions using MTE's. As the next step, we will use the Shenoy-Shafer architecture to propagate in the MTE network and to compute the marginal distribution of the project completion time.

This section describes the operations necessary to carry out propagation in our MTE network example. The operations described here are as follows: Restriction, combination, marginalization, normalization, operations with linear deterministic equations and finding the maximum of two distributions using MTE's. The class of MTE potentials is closed under these operations and this allows us to use the Shenoy-Shafer architecture [Shenoy and Shafer, 1990] to propagate the MTE potentials in the

network. The definitions of restriction, combination, marginalization and normalization are described in Moral *et al*. [2001]. The operations with linear deterministic variables in MTE networks are described in Cobb and Shenoy[2005]. The operations for finding the maximum of two distributions using MTE's are first described here.

3.3.3.1 Restriction

Restriction is the operation of entering evidence during the propagation. In restriction, known variables are substituted with their values.

Let ϕ be an MTE potential for $X = Y \cup Z$. Suppose we receive the evidence for a set of variables $X' = Y' \cup Z' \subseteq X$, s.t. its values $x^{\downarrow \Omega_{x'}}$ are as follows: $x' = (y', z')$. After receiving the evidence the values of the variables are known. Accordingly, the potential ϕ should be updated. The new potential defined on $\Omega_{X \backslash X'}$ is as follows:

$$\phi^{R(X' = x')}(w) = \phi^{R(Y' = y', Z' = z')}(w) = \phi(x) \qquad (3.4)$$

for all $w \in \Omega_{X \backslash X'}$ such that $x \in \Omega_X$, $x^{\downarrow \Omega_{X \backslash X'}} = w$ and $x^{\downarrow \Omega_{X'}} = x'$. In this definition each occurrence of X' in ϕ is replaced with x'. An example for restriction will be provided in section 3.5.

3.3.3.2 Combination

MTE potentials are combined by pointwise multiplication. Let ϕ_1 and ϕ_2 be the MTE potentials for $X_1 = Y_1 \cup Z_1$ and $X_2 = Y_2 \cup Z_2$. The combination of ϕ_1 and ϕ_2 is a new MTE potential for $X = X_1 \cup X_2$ defined as follows:

$$\phi(x) = \phi_1(x^{\downarrow X_1}) \phi_2(x^{\downarrow X_2}) \text{ for all } x \in \Omega_x$$

3.3.3.3   Marginalization

MTE potentials are marginalized by summing over discrete variables and integrating over continuous variables. Let ϕ be an MTE potential for $X = Y \cup Z$. The MTE potentials are closed under marginalization, so the marginal of ϕ for the set of variables $X' = Y' \cup Z' \subseteq X$ is a MTE potential which is computed as follows:

$$\phi^{\downarrow X'}(y', z') = \sum_{y \in \Omega_{Y \backslash Y'}} (\int_{\Omega_{Z \backslash Z'}} \phi(y, z) \, dz'') \qquad (3.5)$$

where $z = (z', z'')$, and $(y', z') \in \Omega_{X''}$. The variables can be marginalized in any sequence, discrete before continuous or continuous before discrete as shown in Formula 3.5.

In the process of marginalization, when the limits of integration include linear functions, then we may end up with linear terms in the remaining variables. These linear terms can be replaced with an MTE approximation so that the result of the marginalization is again an MTE potential. For a linear term x defined over the domain $[x_{min}, x_{max}]$, we replace x with

$$
\begin{aligned}
x_{\min} + (x_{\max} - x_{\min})(0.5*(-13.5070292 + 13.5070292 \, Exp[\frac{0.0726981(x - x_{\min})}{(x_{\max} - x_{\min})}] \\
+ 0.5*(13.5070364 - 13.5070364 \, Exp[\frac{-(0.0754406(x - x_{\min})}{(x_{\max} - x_{\min})}]
\end{aligned}
\qquad (3.6)
$$

The replacement of the linear terms ensures that MTE potentials are closed under marginalization.

3.3.3.4   Normalization

Let $X = Y \cup Z$ be a set of variables where Y is a discrete and Z is a continuous variable. Let ϕ' be the MTE potential for X. Normalization constant for K is calculated as follows:

$$K = \sum_{y \in \Omega_Y} (\int_{\Omega_Z} \phi'(y,z)dz) \quad (3.7)$$

If join trees are initialized with normalized potentials the normalization constant equals to one when no evidence is observed.

3.3.3.5   Linear Deterministic Equations

Consider the PERT Bayes net given in Figure 3.6. The variable C_3 is a deterministic variable containing a linear deterministic equation. As we solve the MTE PERT network using fusion algorithm in Section 3.3.4 each variable is going to be deleted from the network following a deletion sequence. As variables are removed from the network the operations of combination and marginalization will be used, which are explained as above. However, the marginalization operation is different when the variable being deleted is contained in a linear deterministic equation in the network. If it is the case, then we proceed as follows: We solve the equation for the variable being deleted and then substitute this solution in the updated potentials in the network.

Let ψ denote the distribution of $Y|x \sim f_{Y|x}$ and let ζ denote the equation $Z = X + Y$. Suppose we want to delete the variable Y from the network. By solving the

equation for Y and substituting the solution in $f_{Y|x}$ we can remove Y out of the combination and hence find the distribution of $Z|x$. The details are as follows:

$$(\zeta \otimes \psi)^{-Y} = ([Z = X + Y] \otimes f_{Y|x}(y))^{-Y} = ([Y = Z - X] \otimes f_{Y|x}(y))^{-Y} = f_{Y|x}(z - x)$$

3.3.3.6   Maximum of Two Distributions

Finding the distribution of the maximum of two or more distributions has been the interest of many communities of researchers. Especially in the domains of project management, this problem occupies an important place since the completion time of an activity is the sum of its duration and the maximum between the completion times of its immediate predecessors. For this reason, it can be concluded that an accurate estimation of the project completion time is very much affected by an accurate estimation of the activity completion times.

In view of that, in Section 2.3.2, we found the marginal distribution of the maximum of two normally distributed random variables by converting the Bayes net into a MoG Bayesian network and using the Lauritzen-Jensen algorithm to compute the marginals of the MoG Bayesian network.
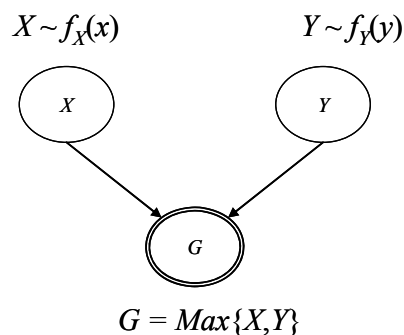
Figure 3.10. Maximum of two distributions

In this section, as we solve the PERT network using MTE's, again we will need to find the maximum of two distributions. The marginal probability density function of the maximum of two distributions can be computed by brute force using order statistics. Consider the small BN given in Figure 3.10. X and Y are continuous variables which have density functions $f_X(x)$ and $f_Y(y)$, respectively. G is a deterministic variable which is distributed as $G = Max\{X, Y\}$. Let F_G denote the cumulative distribution function (CDF) of G, F_X denote the CDF of X and F_Y denote the CDF of Y. Then, $F_G(g) = F_X(g) F_Y(g)$. Therefore, the probability density function of G is given by $f_G(g) = (d/dg)F_G(g) = f_X(g) F_Y(g) + F_X(g) f_Y(g)$, where f_X and f_Y are the PDFs of X and Y, respectively. Since there is no closed form expression for the CDF of a normal distribution, there is no closed form expression for $f_G(g)$ when X and Y are normally distributed. Since MTE potentials are closed under integration both $F_X(g)$ and $F_Y(g)$ can be expressed as MTE potentials. And since MTE potentials are closed under multiplication and addition $f_G(g)$ can also be expressed as MTE potentials. Then, by using the MTE approximations of X and Y, we can obtain an MTE approximation for the distribution of $f_G(g)$.

3.3.4   Fusion Algorithm

The fusion algorithm, first described by Cannings *et al.* [1978], is used to compute the marginal for a variable using local computation [Shenoy, 1992]. Shenoy [1995] described the fusion algorithm as a guide to construct join trees where Shenoy-Shafer architecture will be used to compute the marginals of the variables. The basic idea of

the fusion algorithm is to delete all the variables in the network successively, until we end up with the marginal distribution of the variable of interest.

In this research, we are interested in computing the marginal distribution of the project completion time. Hence, using fusion algorithm, the variables in the MTE PERT Bayes net will be deleted successively, until we end up with the marginal distribution of the project completion time, F. Though different deletion sequences may lead to different computational efforts, the outcome of the network does not get affected with the deletion sequence used. In this example, we will use the deletion sequence D_3, D_1, (D_2, C_3) in order to find the marginal distribution of the project completion time. Figure 3.11 illustrates the construction of the join tree for the PERT example. The messages necessary to compute the marginal distribution of the project completion time are given in Table 3.1 below.
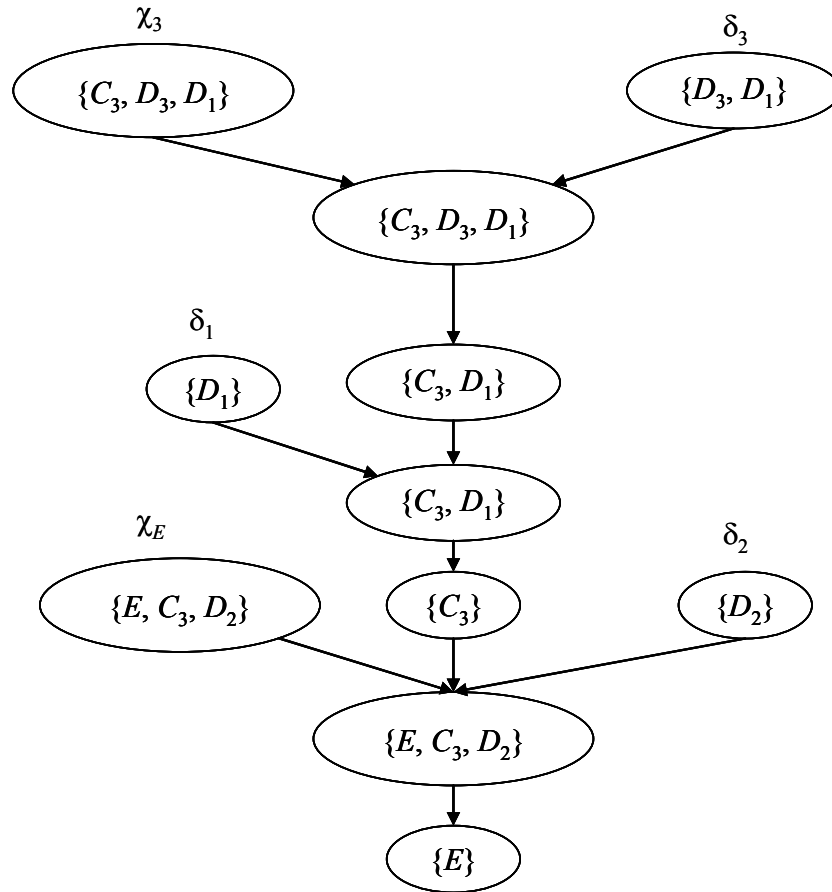
Figure 3.11. Creation of the binary join tree using the fusion algorithm.

Table 3.1. The domains and the potentials of the variables as the binary join tree is

created using the fusion algorithm

Domain	Potential	Distribution
$\{D_1\}$	δ_1	$D_1 \sim N(0.4, 0.01)$
$\{D_2\}$	δ_2	$D_2 \sim N(1.4, 0.01)$
$\{D_3, D_1\}$	δ_3	$D_3 \mid d_1 \sim N(0.6+d_1, 0.04)$
$\{C_3, D_3, D_1\}$	χ_3	$C_3 = D_1 + D_3$
$\{E, C_3, D_2\}$	χ_E	$E = Max\{D_2, C_3\}$

After fusion with respect to D_3

Domain	Potential
$\{D_1\}$	δ_1
$\{D_2\}$	δ_2
$\{C_3, D_1\}$	$(\delta_3 \otimes \chi_3)^{\downarrow\{C_3, D_1\}}$
$\{E, C_3, D_2\}$	χ_E

After fusion with respect to D_1

Domain	Potential
$\{D_2\}$	δ_2
$\{C_3\}$	$((\delta_3 \otimes \chi_3)^{\downarrow\{C_3, D_1\}} \otimes \delta_1)^{\downarrow\{C_3\}}$
$\{E, C_3, D_2\}$	χ_E

After fusion with respect to $\{C_3, D_2\}$

Domain	Potential
$\{E\}$	$((((\delta_3 \otimes \chi_3)^{\downarrow\{C_3, D_1\}} \otimes \delta_1)^{\downarrow\{C_3\}}) \otimes \delta_2 \otimes \chi_E)^{\downarrow\{E\}}$

Fusion with respect to D_3:

Fusion w.r.t. D_3, refers to removing the variable D_3 from the network. This will be done first by combining all the potentials that contain D_3 and next by removing D_3 out of the combination by marginalizing the combination down to the remaining variables. Let $f_{D_3|d_1}$ denote the distribution of $D_3|d_1$. Let χ_3 denote the equation for $C_3 = D_1 + D_3$. By solving the equation for D_3 and substituting D_3 in $f_{D_3|d_1}$ we can find the distribution of $C_3|d_1$. The details are as follows:

$C_3 = D_1 + D_3$

$D_3 = C_3 - D_1$

$$f_{C3|d1}(c_3) = f_{D3|d1}(c_3 - d_1)$$

The probability distribution of $C_3|d_1$ is approximated by an MTE potential as follows:

$$(\delta_3 \otimes \chi_3)^{\downarrow\{C_3, D_1\}}(c_3, d_1) = \begin{cases} 5.0135(-0.010564 + 197.056\exp\{11.2842(-0.6 + c_3 - 2d_1)\} \\ -461.439\exp\{11.7171(-0.6 + c_3 - 2d_1) \\ +264.793\exp\{12.0216(-0.6 + c_3 - 2d_1)\})\} \\ if \quad (d_1 \le c_3 - d_1 < 0.6 + d_1)\,\&\&\,(0.1 \le d_1 < 0.7) \\ \\ 5.0135(-0.010564 + 264.793\exp\{-12.0216(-0.6 + c_3 - 2d_1)\} \\ -461.439\exp\{-11.7171(-0.6 + c_3 - 2d_1) + \\ 197.056\exp\{-11.2842(-0.6 + c_3 - 2d_1)\})\} \\ if \quad (0.6 + d_1 \le c_3 - d_1 \le 1.2 + d_1)\,\&\&\,(0.1 \le d_1 < 0.7) \end{cases}$$

Fusion with respect to D_1:

The variables whose domains contain D_1, (D_1 itself and $C_3|d_1$), are both continuous variables, so deleting D_1 from the network involves finding the joint $f_{C3, D1}(c_3, d_1)$ and integrating this combination over the domain of D_1. The details are as follows:

$$f_{C3, D1}(c_3, d_1) = f_{C3|d1}(c_3) f_{D1}(d_1)$$

$$(f_{C3, D1}(c_3, d_1))^{\downarrow C_3} = \int f_{C3,D1}(c_3, d_1)\, dd_1 = f_{C3}(c_3)$$

Notice that the MTE approximations for both $f_{D1}(d_1)$ and $f_{C3|d1}(c_3)$ are 2 piece three term MTE potentials, so when they are combined the result is an MTE potential with four pieces. In cases, where the domains of functions overlap, the number of pieces can be less than the product of number of pieces. The MTE approximation for $f_{C3, D1}(c_3, d_1)$ is as follows:

$$(\delta_1 \otimes (\delta_3 \otimes \chi_3)^{\downarrow\{C_3,D_1\}})(c_3,d_1)=$$

$$
\begin{cases}
50.2704(-0.010564+197.056\exp\{11.2842(-0.6+c_3-2d_1)\} \\
-461.439\exp\{11.7171(-0.6+c_3-2d_1) \\
+264.793\exp\{12.0216(-0.6+c_3-2d_1)\}\}) \\
(-0.010564+197.056\exp\{22.5684(-0.4+d_1)- \\
461.439\exp\{23.4341(-0.4+d_1)+264.793\exp\{24.0433(-0.4+d_1)\})\} \\
if \quad (d_1 \le c_3-d_1 < 0.6+d_1)\&\&(0.1 \le d_1 < 0.4) \\[2mm]
50.2704(-0.010564+264.793\exp\{-12.0216(-0.6+c_3-2d_1)\} \\
-461.439\exp\{-11.7171(-0.6+c_3-2d_1)+ \\
197.056\exp\{-11.2842(-0.6+c_3-2d_1)\})(-0.010564+197.056\exp\{22.5684(-0.4+d_1)- \\
461.439\exp\{23.4341(-0.4+d_1)+264.793\exp\{24.0433(-0.4+d_1)\}\})\} \\
if \quad (0.6+d_1 \le c_3-d_1 \le 1.2+d_1)\&\&(0.1 \le d_1 < 0.4) \\[2mm]
50.2704(-0.010564+197.056\exp\{11.2842(-0.6+c_3-2d_1)\} \\
-461.439\exp\{11.7171(-0.6+c_3-2d_1) \\
+264.793\exp\{12.0216(-0.6+c_3-2d_1)\}\}) \\
(-0.010564+264.793\exp\{-24.0433(-0.4+d_1)- \\
461.439\exp\{-23.4341(-0.4+d_1)+197.056\exp\{-22.5684(-0.4+d_1)\})\} \\
if \quad (d_1 \le c_3-d_1 < 0.6+d_1)\&\&(0.4 \le d_1 \le 0.7) \\[2mm]
50.2704(-0.010564+264.793\exp\{-12.0216(-0.6+c_3-2d_1)\} \\
-461.439\exp\{-11.7171(-0.6+c_3-2d_1)+ \\
197.056\exp\{-11.2842(-0.6+c_3-2d_1)\})(-0.010564+264.793\exp\{-24.0433(-0.4+d_1)- \\
461.439\exp\{-23.4341(-0.4+d_1)+197.056\exp\{-22.5684(-0.4+d_1)\})\} \\
if \quad (d_1 \le c_3-d_1 < 0.6+d_1)\&\&(0.4 \le d_1 \le 0.7)
\end{cases}
$$

The expected value and variance for the marginal of C_3 are calculated as 1.4 and 0.0786. These answers are comparable with results from multivariate normal theory, which gives an expected value and variance of 1.4 and 0.08. The plots of the MTE approximations for $f_{C_3, D_1}(c_3, d_1)$ and $f_{C_3}(c_3)$ are given in Figure 3.12 and Figure 3.13, respectively.
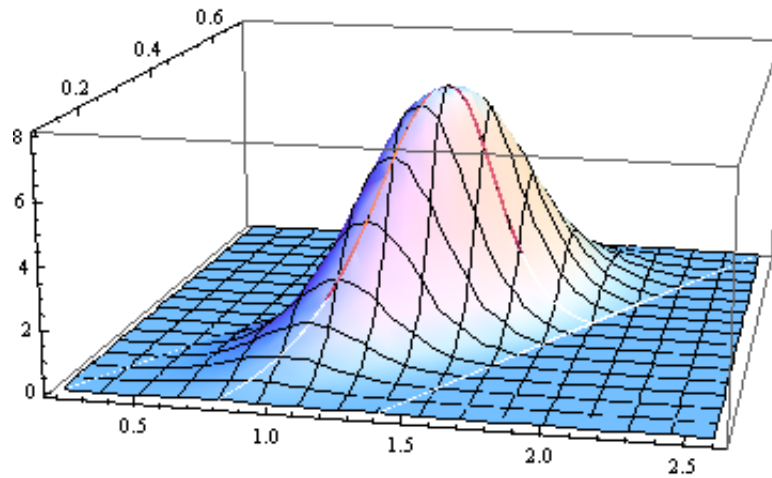
Figure 3.12. The MTE approximation for $f_{C_3, D_1}(c_3, d_1)$
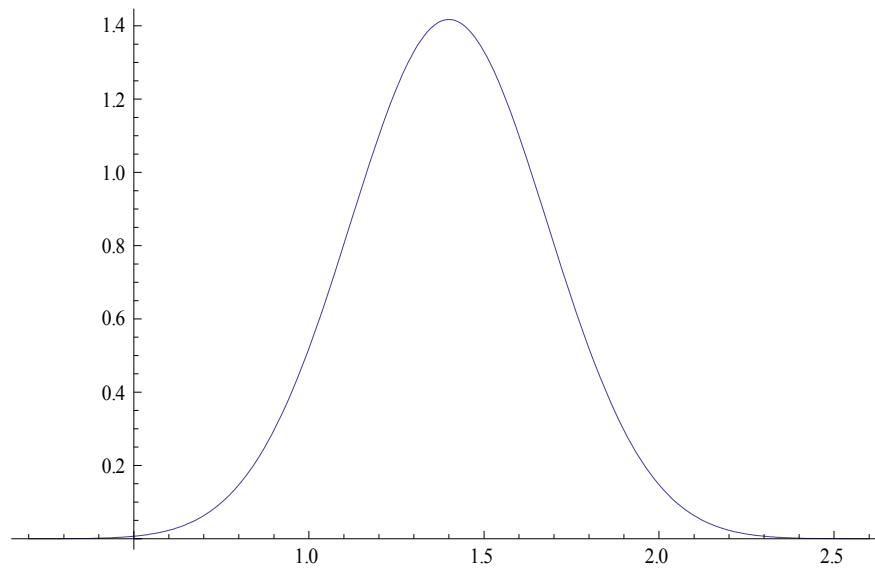
Figure 3.13. The plot of the approximation for $f_{C_3}(c_3)$

The next step is to find the marginal distribution of $E = Max\{C_3, D_2\}$ which requires the variables, C_3 and D_2, to be deleted at the same time.

Figure 3.14 represents the current state of our network after the variables D_3 and D_1 are removed from the network. As the next and final step, we have to find the project completion time $E = Max\{C_3, D_2\}$ which requires the variables C_3 and D_2 to be deleted at the same time.

$$C_3 \sim f_{C_3}(c_3) \qquad D_2 \sim N(1.4, 0.01)$$

$$E = Max\{C_3, D_2\}$$
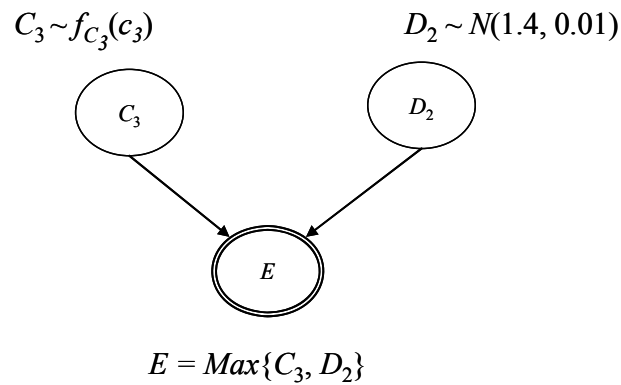
Figure 3.14. The conditional distribution of E after D_3 and D_1 are deleted from the

network

As shown before, the probability density function of F_E is given by $f_E(e) = (d/de)F_E(e) = f_{C_3}(e) \, F_{D_2}(e) + F_{C_3}(e) \, f_{D_2}(e)$, where f_{C_3} and f_{D_2} are the PDFs of C_3 and D_2, respectively. In sections 3.3.3 and 3.3.4 the PDF's of D_2 and C_3 are approximated as δ_2 and $((\delta_3 \otimes \chi_3)^{\downarrow \{C_3, D_1\}} \otimes \delta_1)^{\downarrow \{C_3\}}$ respectively. As the next step, the CDF of these potentials are calculated to find the MTE approximation of $f_E(e)$. The MTE approximation for $F_{D_2}(e)$ is calculated as follows:

$$\left\{ \begin{array}{l} \displaystyle\int_{1.1}^{e} (10.027(-0.010564+197.056\exp\{22.5684(-1.4+e)-461.439 \\ \exp\{23.4341(-1.4+e)+264.793\exp\{24.0433(-1.4+e)\})\})de \quad if \;\; 1.1 \leq e < 1.4 \\[2em] \displaystyle\int_{1.1}^{1.4} (10.027(-0.010564+197.056\exp\{22.5684(-1.4+e)-461.439 \\ \exp\{23.4341(-1.4+e)+264.793\exp\{24.0433(-1.4+e)\})\})de + \\ \displaystyle\int_{1.4}^{e} (10.027(-0.010564+264.793\exp\{-24.0433(-1.4+e)-461.439 \\ \exp\{-23.4341(-1.4+e)+197.056\exp\{-22.5684(-1.4+e)\})\})de \; if \;\; 1.4 \leq e \leq 1.7 \\[2em] \displaystyle\int_{1.1}^{1.4} (10.027(-0.010564+197.056\exp\{22.5684(-1.4+e)-461.439 \\ \exp\{23.4341(-1.4+e)+264.793\exp\{24.0433(-1.4+e)\})\})de + \\ \displaystyle\int_{1.4}^{1.7} (10.027(-0.010564+264.793\exp\{-24.0433(-1.4+e)-461.439 \\ \exp\{-23.4341(-1.4+e)+197.056\exp\{-22.5684(-1.4+e)\})\})de \; if \;\; e > 1.7 \end{array} \right.$$

The plot for the approximation for the CDF of D_2 is illustrated in Figure 3.15 below.
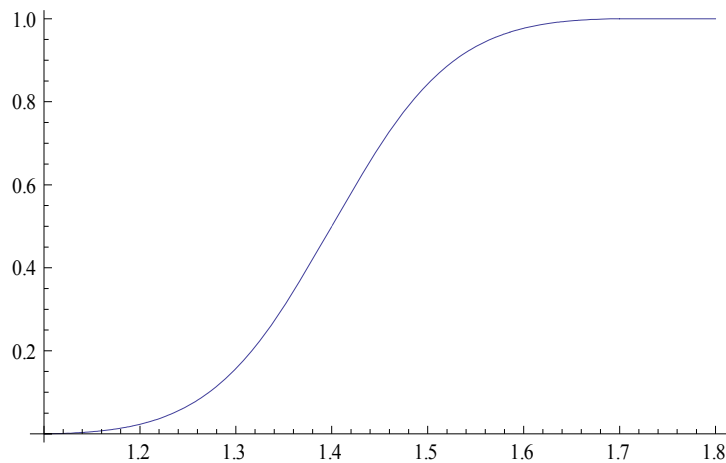
Figure 3.15. Approximation for $F_{D_2}(e)$

The approximation for $F_{C_3}(e)$ is approximated using the following potential:

$$
\begin{cases}
\begin{aligned}
&0.0000365286 + 0.0497112\exp\{112842e\} - 0.0000493421\exp\{11.2842e\} + \\
&0.0538651\exp\{11.7171e\} - 0.0439047\exp\{11.7171e\} - 0.0000323852\exp\{12.0216e\} \\
&-0.0493977\exp\{12.0216e\} + \exp\{12.0216e\}(-0.0288732 + 0.00719167e) \\
&+\exp\{11.2842e\}(0.0183284 + 0.0119137e) + \\
&\exp\{11.7171e\}(0.000318218 + 0.0343224e) + 0.000194132e + 0.00140252e^2 \\
&\quad if \qquad 1/5 \le e < 4/5
\end{aligned} \\[6pt]
\begin{aligned}
&0.0159207 + \exp\{-35.0229e\}(-4344.14\exp\{23.0013e\} + 6409.05\exp\{23.3059e\} + \\
&2165.63\exp\{23.7387e\} - \\
&0.0644511\exp\{46.3071e\} + 0.0338628\exp\{46.74e\} + \\
&\exp\{46.74e\}(-0.0444922 - 0.0343224e) \\
&+\exp\{46.3071e\}(-0.00329702 - 0.0119137e) + \exp\{47.0445e\}(0.0784474 - 0.00719167e) \\
&+0.00140252\exp\{35.0229e\}(-80.7328 + e)(-0.530627 + e)) \\
&\quad if \qquad 4/5 \le e < 7/5
\end{aligned} \\[6pt]
\begin{aligned}
&0.5 + \exp\{-35.0229e(4.10866\times10^{-11}\exp\{46.3071e\} - 3.61879\times10^{-11}\exp\{46.74e\} + \\
&1.04544\times10^{-11}\exp\{47.0445e\} + \exp\{23.3059e\}(1.89028\times10^{13} - 6.07866\times10^{12}e) + \\
&\exp\{23.0013e\}(-2.42299\times10^{13} - 2.98836\times10^{12}e) + \exp\{23.7387e\}(5.32922\times10^{12} - 6.27959\times10^{11}e) \\
&-0.00140252\exp\{35.0229e\}(-.36384 + e)(82.0272 + e)) \\
&\quad if \qquad 7/5 \le e < 2
\end{aligned} \\[6pt]
\begin{aligned}
&0.980479 + \exp\{-35.0229e\}(-0.00140252\exp\{35.0229e\}(-6.23525 + e)(0.496835 + e) \\
&+\exp\{23.7387e\}(-5.34198\times10^{12} + 6.27959\times10^{11}e) + \exp\{23.0013e\} \\
&(2.417\times10^{13} + 2.98836\times10^{12}e) + \exp\{23.3059e\}(-1.88406\times10^{13} + 6.07866\times10^{12}e)) \\
&\quad if \qquad 2 \le e < 13/5
\end{aligned} \\[6pt]
\begin{aligned}
&1 \\
&\quad if \quad e \ge 13/5
\end{aligned}
\end{cases}
$$

The approximation for $F_{C_3}(e)$ is represented in Figure 3.16 below.
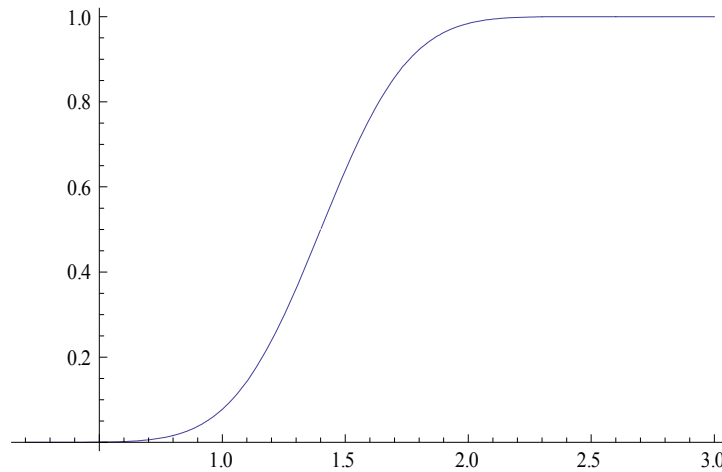
Figure 3.16. MTE Approximation for $F_{C_3}(e)$

Notice that the approximations for the PDF and CDF of D_2 have three and four pieces, respectively. With four pieces for the approximation of $f_{C_3}(e)$ and five pieces for $F_{C_3}(f)$, the number of pieces required for the approximation of $f_E(e)$ can be up to 12 pieces. However, in this instance the domains do overlap. As a result, we have 4 pieces for the approximation of $f_E(e)$.

Notice that both of the approximations for C_3 and for the cdf of D_2 have linear terms in the potentials. The approximation for $f_E(e)$ is approximated using the following potential:

$2.57799 \times 10^{-43} \exp\{-105.069e\} (2.39444 \times 10^{46} \exp\{93.0471e - 3.44977 \times 10^{46}$

$\exp\{93.3517e\} + 1.12591 \times 10^{46} \exp\{93.7845e\} - 7.95894 \times 10^{40} \exp\{105.069e\}$

$-2.952 \times 10^{35} \exp\{115.616e\} + 4.48096 \times 10^{35} \exp\{115.92e\} - 2.82116 \times 10^{41}$

$\exp\{116.353e\} - 1.94871 \times 10^{40} \exp\{116.353e\} + 2.14392 \times 10^{35} \exp\{116.481e\}$

$-2.17488 \times 10^{41} \exp\{116.786e\} + 1.54444 \times 10^{41} \exp\{116.786e\} +$

$3.64876 \times 10^{41} \exp\{117.09e\} + 1.13784 \times 10^{35} \exp\{117.219e\} +$

$+8.14373 \times 10^{34} \exp\{117.395e\} - 2.84842 \times 10^{34} \exp\{117.828e\} +$

$1.03173 \times 10^{31} \exp\{127.637e\} - 7.20917 \times 10^{30} \exp\{128.503e\} +$

$1.76627 \times 10^{30} \exp\{129.112e\} - 1.48536 \times 10^{31} \exp\{138.921e\} - 2.56923 \times 10^{30}$

$\exp\{139.354e\} + 1.74369 \times 10^{31} \exp\{139.659e\} + 1.02229 \times 10^{31} \exp\{139.787e\}$

$+1.76418 \times 10^{30} \exp\{140.22e\} - 2.47951 \times 10^{30} \exp\{140.396e\} - 1.19968 \times 10^{31}$

$\exp\{140.524e\} - 4.27225 \times 10^{29} \exp\{140.829e\} + 2.90909 \times 10^{30} \exp\{141.134e\}$

$-2.14578 \times 10^{46} \exp\{93.0471e\}e + 3.08554 \times 10^{46} \exp\{93.3517e\}e - 1.0041 \times 10^{46}$

$\exp\{93.7845e\}e + 9.48498 \times 10^{40} \exp\{105.069e\}e + 2.98827 \times 10^{41} \exp\{116.353e\}e$

$-3.1967 \times 10^{40} \exp\{116.353e\}e + 7.17633 \times 10^{40} \exp\{116.786e\}e - 1.63027 \times 10^{41}$

$\exp\{116.786e\}e - 4.18265 \times 10^{41} \exp\{117.09e\}e - 1.65548 \times 10^{31} \exp\{127.637e\}e$

$+1.15377 \times 10^{31} \exp\{128.503e\}e - 2.82197 \times 10^{30} \exp\{129.112e\}e - 2.59855 \times 10^{30}$

$\exp\{138.921e\}e - 7.58194 \times 10^{30} \exp\{139.354e\}e - 1.60278 \times 10^{30} \exp\{139.659e\}e$

$+1.78867 \times 10^{30} \exp\{139.787e\}e + 5.21725 \times 10^{30} \exp\{140.22e\}e - 4.33869 \times 10^{29}$

$\exp\{140.396e\}e + 1.10266 \times 10^{30} \exp\{140.524e\}e - 1.26526 \times 10^{30} \exp\{140.829e\}e$

$-2.67371 \times 10^{29} \exp\{141.134e\}e - 1.72881 \times 10^{39} \exp\{105.069e\}e^2 + 5.52377 \times 10^{40}$

$\exp\{116.353e\}e^2 + 1.6524 \times 10^{41} \exp\{116.786e\}e^2 + 3.55232 \times 10^{40} \exp\{117.09e\}e^2$

$+2.0394 \times 10^{29} \exp\{127.637e\}e^2 - 1.42129 \times 10^{29} \exp\{128.503e\}e^2 + 3.47618 \times 10^{28}$

$\exp\{129.112e\}e^2)$

if $\quad 11/10 \leq e < 7/5$

$$\exp\{-221.855e(-3.9961\times10^{31}e^{185.79e}+3.1297\times10^{31}\exp\{186.094e\}+2.99359\times10^{31}$$
$$\exp\{186.399e\}+8.66782\times10^{30}\exp\{186.527e\}-2.3447\times10^{31}\exp\{186.703e\}$$
$$-6.49175\times10^{30}\exp\{187.136e\}-3.85386\times10^{30}\exp\{187.264e\}+3.0188\times10^{30}$$
$$\exp\{187.569e\}+8.35428\times10^{29}\exp\{188.002e\}+1.36423\times10^{18}\exp\{197.811e\}$$
$$-1.01336\times10^{18}\exp\{198.42e\}+1.28811\times10^{17}\exp\{199.286e\}$$
$$+2.4055\times10^{7}\exp\{209.095e\}-2.04682\times10^{7}\exp\{209.528e\}$$
$$-1.74558\times10^{7}\exp\{209.705e\}+3.44831\times10^{14}\exp\{209.833e\}-2.72166\times10^{14}$$
$$\exp\{210.137e\}-4.17204\times10^{6}\exp\{210.442e\}-7.27037\times10^{13}\exp\{210.57e\}$$
$$-1.81214\times10^{6}\exp\{211.003e\}+5088021\exp\{211.308e\}-0.256498\exp\{221.855e\}$$
$$+6.46071\times10^{-10}\exp\{233.139e\}-4.99559\times10^{-10}\exp\{233.572e\}+1.48099\times10^{-10}$$
$$\exp\{233.876e\}-4.94544\times10^{30}\exp\{185.79e\}e-9.97463\times10^{30}\exp\{186.094e\}e$$
$$+3.70499\times10^{30}\exp\{186.399e\}e-1.01796\times10^{30}\exp\{186.527e\}e+7.47163\times10^{30}$$
$$\exp\{186.703e\}e+7.62356\times10^{29}\exp\{187.136e\}e-4.77011\times10^{29}\exp\{187.264e\}e$$
$$-9.61749\times10^{29}\exp\{187.569e\}e-9.80996\times10^{28}\exp\{188.002e\}e-1.16949\times10^{17}$$
$$\exp\{197.811e\}e+8.68602\times10^{16}\exp\{198.42e\}e-1.1039\times10^{16}\exp\{199.286e\}e$$
$$+1.24291\times10^{13}\exp\{209.833e\}e+1.09306\times10^{14}\exp\{210.137e\}e+1.49155\times10^{13}$$
$$\exp\{210.57e\}e+0.0191513\exp\{221.855e\}e-4.91101\times10^{-11}\exp\{233.139e\}e$$
$$+4.49139\times10^{-11}\exp\{233.572e\}e-1.33126\times10^{-11}\exp\{233.876e\}e-1.54735\times10^{15}$$
$$\exp\{197.811e\}e^{2}+1.14928\times10^{15}\exp\{198.42e\}e^{2}-1.46068\times10^{14}\exp\{199.286e\}e^{2}$$
$$-3.80536\times10^{12}\exp\{209.833e\}e^{2}-7.54442\times10^{12}\exp\{210.137e\}e^{2}-7.50588\times10^{11}$$
$$\exp\{210.57e\}e^{2}+0.000445686\exp\{221.855e\}e^{2}$$
$$if\quad 7/5\le e\le17/10$$

$$\exp\{-116.786e\}(2.88295\times10^{14}\exp\{104.764e-2.27564\times10^{14}\exp\{105.069e\}$$
$$-6.0764\times10^{13}\exp\{105.502e\}-0.106119\exp\{116.786e\}+4.6363\times10^{-10}$$
$$\exp\{128.07e\}-4.24015\times10^{-10}\exp\{128.503e\}+1.25679\times10^{-10}\exp\{128.807e\}$$
$$+3.5925\times10^{13}\exp\{104.764e\}e+7.1224\times10^{13}\exp\{105.069e\}e+7.08602\times10^{12}$$
$$\exp\{105.502e\}e-0.00280504\exp\{116.786e\}e)$$

if $\quad 17/10 < e < 2$

$$\exp\{-46.74e\}(-2.87575\times10^{14}\exp\{34.7183e\}+2.26835\times10^{14}\exp\{35.0229e\}$$
$$+6.09081\times10^{13}\exp\{35.4558e\}+0.00804824\exp\{46.74e\}-3.5925\times10^{13}$$
$$\exp\{34.7183e\}e-7.1224\times10^{13}\exp\{35.0229e\}e-7.08602\times10^{12}\exp\{35.4558e\}e$$
$$-0.00280504\exp\{46.74e\}e$$

if $\quad 2 \le e \le 13/5$

By comparing the means and variances of the approximation with the exact analytic results calculated with Clark's method [1961] we can evaluate the goodness of our approximation. Accordingly, the mean of the marginal distribution of E is calculated as 1.51883 and 0.0300638, respectively. Comparing to 1.51968 and 0.0306761 given by the exact analytic results, the approximation is quite good. The approximation for $f_E(e)$ overlaid on the actual distribution is given in Figure 3.17 below.
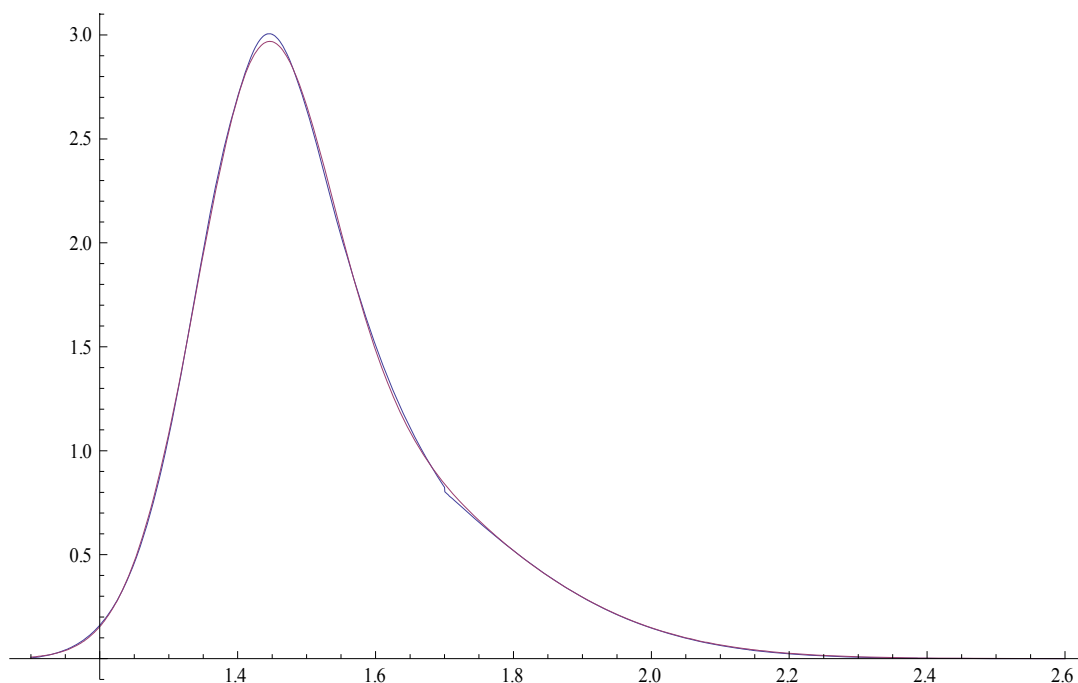
Figure 3.17. Approximation of $f_E(e)$ overlaid on the actual distribution

After normalization, when the limits of integration include linear terms, then we may end up with linear terms in the remaining variables as it is the case with the approximation of C_3 and of the cdf of D_2. These linear terms can be approximated again using MTE potentials, such that it can ensured that the result is again an MTE approximation and MTE's are closed under marginalization. However, replacing the linear terms with the MTE potentials causes bad accuracy in our approximations.

3.5   Entering Evidence in a MTE PERT Network

In this research first we transformed a PERT network into a PERT Bayesian network which we then solved using mixtures of truncated exponentials. We evaluated our results for the mean and variance of the project completion time with the exact analytic results calculated with Clark's method [1961] and we evaluated the shape of

the marginal distribution of the project completion time by comparing to the actual distribution. In this context, it is natural to question our methods described in this research and ask for the advantage we obtain by using the methods described, instead of using straight forward simulation methods that are already handy.

With simulation methods the activity durations can be represented realistically. As it is the case with our methods, the activity durations can have any type of distribution and one can also represent the correlation between the activity durations. However, with straight-forward Monte Carlo simulation methods we can not include the observations of continuous variables and update our inferences accordingly. On the other hand, by transforming the PERT network into a Bayesian network and solving it using the Shenoy-Shafer architecture we can update our network, once evidence is observed, and find the posterior distributions of the activities and get a more accurate estimation for the project completion time.

Consider the PERT Bayesian network given in Figure 3.18. This is a PERT Bayes net with four activities A_1, A_2, A_3 and A_4. Suppose we know that the activities A_1 and A_2 will be performed by the same contractor. The quality of the work done by this contractor is distributed as $f_Q(q)$. The quality of the work performed by the contractor effects the duration of the activities A_1 and A_2 such that with higher quality it will take less time to complete these activities. In addition to these, we also have the information that the same contractor performs another activity similar to ours within the firm. This activity A_4 is outside of our project but we included it in our network in

Figure 3.18 anyway since it will effect our later conclusions. As you can see in Figure 3.18 the duration of activity A_4 also depends on the quality of the contractor's job.
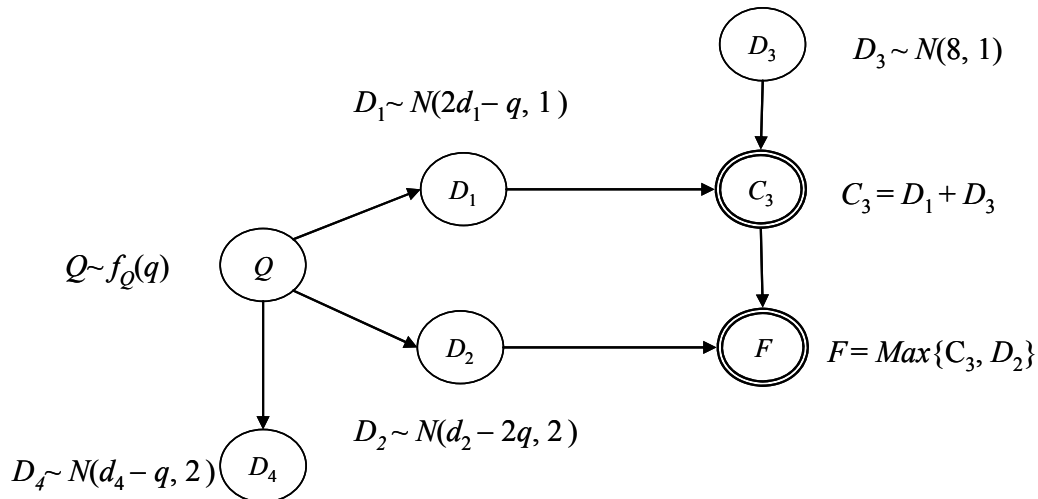
Figure 3.18. Representation of the example as a PERT Bayesian network

This example given in Figure 3.18 can be solved using the means of simulation methods as well as with the methods represented throughout this research. However, suppose we observe that the duration of activity A_4 lasted 10 days to complete. Hence we have the evidence $e_{D_4} = 10$. With the methods described in this dissertation this evidence can be incorporated in the network and the estimates for the durations can be updated accordingly, which is not possible using the straight forward simulation. With our method we can find the posterior distribution of Q after receiving the evidence e_{D_4} which in turn will change the estimates for the distributions of A_1 and A_2 and consequently the estimate for the project completion time. Including the observations in the network and updating the distributions accordingly will improve the quality of the inference. The PERT BN after receiving the evidence e_{D_4} is represented in Figure 3.19 below.
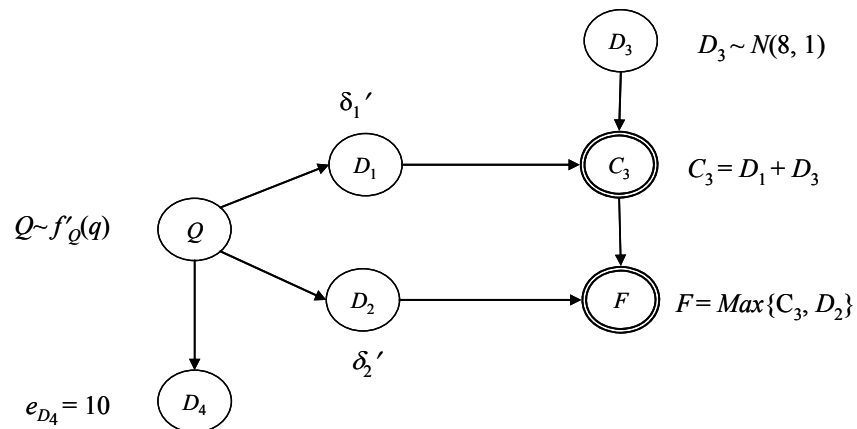
Figure 3.19. The PERT Bayesian network after receiving the evidence e_{D_4}

3.6   Summary and Conclusions

In Chapters 1 and 2 we illustrated how a PERT network can be transformed into a PERT Bayesian network and be solved using MoG's. This method is promising in the sense that it eliminates the false assumptions made in the literature which overlook the true distribution of the maximum of two distributions and thus fail to compute an accurate estimation for the project completion time. However, we also concluded that the restricted nature of MoG Bayesian networks can make the inference process cumbersome. With this chapter we provide a different method, an alternative to MoG Bayesian networks, which overcomes the difficulties involved in solving stochastic PERT networks using MoG's, but still possess the advantages involved in it.

In this research we transformed a PERT network first into a PERT Bayesian network which allows us to depict the dependencies between the activities. By transforming it into a Mixtures of Truncated exponentials network as the next step, we were able to use the Shenoy-Shafer architecture to propagate the MTE potentials

and thus to find the marginal distribution of the project completion time. Finding the distribution of the project completion time is important because there is no closed form expression for the distribution of the maximum of two normal distributions and this fact, previously forced the researchers to make false assumptions like the maximum of two normal distributions is again normally distributed. However, in this research we showed that by approximating the maximum of two distributions using MTE's a very accurate estimation for the project completion time can be obtained. Comparing our results for mean and variance with the exact analytic results using Clark's method [1961] our method proved to be successful. We evaluated the shape of our distribution by comparing to the actual distribution calculated by brute force using order statistics. Also in this case our approximation proved to be successful.

In this research both MoG's and MTE's are provided as tools to make inferences in stochastic PERT networks. With their ability to find accurate estimations for the true distributions of the maximum of two distributions and for the project completion time each can be considered as helpful in achieving our goals. However it should be noted that the inference process using MTE PERT networks, compared to MoG's, is more straightforward in the sense that the MTE PERT networks do not force restrictive settings like, the inability of discrete variables to have continuous parents as it is the case with MoG networks.

Comparing our method to straight forward simulation on the other hand, the MTE PERT Bayesian networks possess the advantage that the observations can be integrated to the inference process. Once evidence is observed we can update our

network accordingly and find the posterior distributions of the activities and thus obtain a more accurate estimation for the project completion time.

The drawback with our method is on the other hand, that the number of exponential terms increases rapidly as the fusion algorithm is applied which in turn makes the inference process more difficult to apply. Additionally, in the process of marginalization, when the limits of integration include linear functions, we may end up with linear terms in the remaining variables. These linear terms can be approximated using an MTE approximation and it can be ensured that the result is again an MTE potential. However, replacing the linear terms with the MTE potentials causes bad accuracy in our approximations.

4 USING RADIO FREQUENCY IDENTIFICATION IN OPERATIONS MANAGEMENT

4.1 Introduction

Radio Frequency Identification (RFID) is a generic term for a variety of technologies that use radio waves to automatically identify individual items [Cavoukian, 2004]. This technology known for over 50 years, prepares to have its real bang in the business world after its potential for commercial applications has been realized. The capability of identifying individual products, ability to track the products through the processes, differentiates RFID from its preceding alternatives; but the real and huge potential of RFID systems is hidden in the massive amount of data that is captured by RFID systems. The following subsections discuss the current state of grocery shopping and the capabilities of RFID technology in the domain of operations management.

4.2 Grocery Shopping

Love it or hate it, grocery shopping occupies a significant amount of time of your life. It may seem as a straightforward task—all you need is just a shopping list. However, almost 60% of household supermarket purchases are unplanned and the result of in store decisions [Inman and Winer, 1999]. Even having a shopping list is sometimes not enough. The huge variety of products offered turns the grocery stores into labyrinths, so you have to be cautious not to get lost between the aisles as you search

for the products on your list. By the time you find the item you are looking for, you may be overwhelmed to see how many different brands offer the same item.

Grocery basket selection can be thought as a reflection of customers' needs. Ideally, the products selected should represent the results of a comparison made by the customer based on the price and quality aspects of the products. Considering the nature of a simple grocery-shopping trip described earlier, a careful selection of products requires the devotion of a significant amount of time and energy on the customers' side. On the contrary, modern life imposes time constraints on the customers, which make them unwilling to spend any more time for grocery shopping than is necessary. As a result, the explosion of the size of product assortments (more than 100,000 references in a large hypermarket) no longer allows for a clear identification of differences in quality and prices inside the product mix [Bruno and Pache, 2005].

The situation on the retailers' side is also not very promising. The competition between the grocery stores is increasing every day, forcing the retailers to find new ways to influence the purchase decisions of the customers. Today a wide variety of methods to track and analyze the customers' behavior in e-commerce systems is available. For instance, amazon.com makes real-time recommendations (Customers who bought this item also bought…) to its customers based on the information of the products that have been put in a shopping cart or reviewed by the customer. However, in traditional retail stores, such systems are not used, and, therefore, the customer's behavior is considered as a black box [Decker, 2003].

As a way to affect the consumers' purchase decisions and to introduce new products, the shelf configurations of the stores are periodically rearranged. Although this might help the retailers to find the optimal allocation of the products, it bothers the customers for not being able to find the products they are looking for.

Another way to influence the customers' purchase decisions is to do promotions. Every day hundreds of items inside the grocery stores are advertised as an effort to trigger the demand of customers for those products on promotion. Whether these special offers will become a subject of interest to customers primarily depends on whether the customers are aware of them or not. Studies suggest that more than half of the shoppers who purchased an item that was on sale were unaware that the price was reduced [Mittal, 1994]. To inform the customers about your ongoing promotions you may increase the rate of your advertisements, which will increase your costs significantly.

Advertisements and promotions are two effective ways to influence sales. However, an advertisement of a promotion will be more successful, if the promotion is particularly advertised to those shoppers who are likely to be interested in the offer. Clearly, the purchase of a product on promotion by an informed customer does not only show the success of the promotion; it is also a valid indicator of customers' interest in that particular product. Thus, the success of a promotion secondarily depends on whether a customer is interested in that product or not.

In order to understand the underlying patterns of customers' purchase decisions, most grocery stores identify its customers through customer loyalty cards via which

they keep track of the products purchased by the customer. Based on this information they tailor promotions to individual customers by giving discount coupons at the checkout. However, these promotions happen after the shopping is over which tremendously reduces the impact of the promotion on the sale. The companies are mostly incapable of influencing the customers' decision making process when they are shopping since the data about the customers' shopping behavior is only available after the decisions are made, i.e., after the shopping is over.

As a way to interact with the customer during the shopping, grocery stores install kiosks from which customers can get information about the ongoing promotions and the products displayed. However, stress and time pressure potentially force a customer to fully concentrate on the original task where the customer is not willing or able to learn the operation of a complex shopping support system [Schneider, 2004].

Having considered all of this, the e-commerce seem to have a huge advantage over traditional grocery shopping because of their capability to make targeted advertising at the same time as the consumer is shopping. Inspired by the real-time recommendation systems of e-commerce we should be looking for ways to transfer the methods of e-commerce systems to the current state of grocery shopping. The capability of RFID technology to identify individual products and collect real time data about the customer behavior inside a store makes a new model for the traditional grocery shopping feasible.

The following section describes the RFID technology and its capabilities in the domain of operations management.

4.3   Radio Frequency Identification

An RFID system consists of two basic parts: a tag and a reader. Readers, depending upon design and technology used, may be a read-only or a read-write device [Finkenzeller, 1999]. They capture the information stored or gathered by the tag. The RFID tags can be either active or passive, depending whether they have their own power supply or not. Active RFID tags offer superior performance. Because they are connected to their own battery, they can be read at a much higher range-from several kilometers away. However, they are larger and more expensive. Passive tags have no power source and no on-tag transmitter, which gives them a range of less than 10-meters and makes them sensitive to environmental constraints [Cavoukian, 2004].

Among the automatic identification systems, barcode technology has been the leader for over 20 years. Nevertheless, with the decreasing cost of the RFID tags, companies have begun to favor RFID systems over barcode technology. Although it is a fact that the reduced costs of the RFID tags have contributed a lot to the present popularity of the RFID systems, this is not the main motive why the RFID systems are preferred over barcodes. In Table 4.1, we illustrate the potential benefits that companies may achieve in their operation management activities by using RFID systems instead of the barcodes.

Table 4.1: The potential benefits of RFID systems in operations management

activities.

	Barcode	RFID	Potential benefit of RFIDs
Data capturing capacity	A barcode can hold only around 1000 characters of data. [Mital, 2003]	Up to 128,000 characters in an RFID chip [Mital, 2003].	The superior data capturing capacity of RFID systems offers enough room for a unique serial number, expiration date or other pertinent information [Sweeney, 2005] -This is a serious drawback of bar codes compared to the RFID systems since in RFID systems information is specific to that individual item [Cavoukian, 2004].
Cost	The barcode system is still a much cheaper identification system than the RFID technology and the experts predict that it will remain to be so.	Today Passive RFIDs sell for less than 50 cents in high volumes, and analysts predict they'll sell for five cents in high volumes by the end of this decade [Dipert, 2005].	Tags are reusable and have very long lives, so in supply chain operations where containers are continually reused, there would be no need to re-label the containers, saving on manpower and other costs associated with label production and fixing [Hopwood, 2005].
Processing times	Only one item can be read at a time because of the line of sight technology required. -The existence of dirt or dust can avoid the reading barcodes.	RFID tags can be read in harsh environments such as snow, fog, etc. with a reading distance ranging from 50 feet to 100 meters and beyond [Cavoukian, 2004]	The processing times of items increases significantly, when bar code systems are in use.
The query of components and subassemblies	Requires positioning the cases so that the labels can be read by the scanners -line of sight reading is required	Automatic check that all items from the bill of material -are received -are placed in the right location -RFID does not require positioning the cases	Convenience in order processing -helps to decrease the labor costs -reduce the order preparation times [Rutner, 2004]
A valid source of information in order preparation and processing	not applicable	How much time a worker spends on the preparation of a particular item can be measured	Management could use this data for -setting benchmarks -evaluating employees -planning labor requirements [Rutner, 2004]

Prevention of Spoilage	not applicable	Sensor-equipped tags can monitor the environment surrounding perishable items and maintain a history of environmental changes	RFID systems can be used -to detect potential spoilage conditions [Curtin, 2005] -to identify the causes of spoilage
Prevention of Theft	not applicable	The capability to locate every individual product within the inventory	Provides a tremendous opportunity for companies to prevent theft
Prevention of Shrinkage	Real time data is not available	Automatic collection of real time data	-the automatic collection of real time data prevents the shrinkage problem, and if not, makes the data available to detect the cause of shrinkage -better replenishments decisions can be made since accurate data are readily available with RFID [Lee, 2004]
Prevention of Stockouts	Captures information on how much is sold form each product	Captures information about the real time data of the current inventory (how much is sold, how much is missing)	The ability of RFID systems to prevent and detect when theft and/or shrinkage is present, makes the data more accurate thus preventing the occurrence of stockouts

4.4　Summary and Conclusions

The use of RFID systems in commercial applications is an emerging trend and RFID is ready to place itself as the dominant technology used in real word applications. However, the importance of RFID technology is not just limited by the convenience it provides. More importantly, RFID systems create massive amounts of data, which gives the ability to track and trace materials at the case-level within the supply chain, and at the item level from manufacturing to post sales. Therefore, the real question to be answered is: How we can transform this massive amount of data into managerially useful information?

5 USING RFIDs AND COLLABORATIVE FILTERING FOR TARGETED ADVERTISING

5.1 Introduction

As mentioned earlier the real potential of RFID systems is hidden in the massive amount of data collected through RFID. This application is a perfect illustration of that. We can use the RFID technology for getting real time information about the consumer behavior as they are shopping and that may enable us to inform the customer about the promotions in store in which the customer is likely to be interested. Using RFID we can get information about the products a customer is placing in his shopping basket, and using collaborative filtering we can advertise those products on promotion in which the customer is more likely to be interested based on what is already in the customer's shopping basket.

5.2 Collaborative Filtering

Collaborative filtering, first introduced by Resnick *et al.* (1994), is defined as predicting preferences of an active user given a database of preferences of other users [Mild, 2002]. Depending on the technology used, recommendation systems are classified in two classes, content-based filtering (CBF) and collaborative filtering (CF). Content-based methods make recommendations by analyzing the description of the items that have been rated by the user and the description of items to be recommended [Pazzani, 1999]. The main difference between collaborative filtering

and content-based filtering is that CF does not rely on the content descriptions of the items, but depends purely on preferences expressed by a set of users [Yu *et al*, 2004]. Since collaborative filtering does not depend on error-prone machine analysis of content, it has significant advantages over traditional content-based filtering (ability to filter any type of content, etc.) [Herlocker *et al.*, 2000].

In e-commerce, collaborative filtering is widely used as a tool for targeted advertising. Using the capabilities of RFID, we might be able to transfer this method to traditional retail stores and base the advertisements on real-time data.

The technique used in collaborative filtering is based either on explicit or implicit voting. The data sets in explicit voting contain users explicit preference ratings for products. Implicit voting refers to interpreting user behavior or selections to impute a vote or preference [Breese *et al.,* 1998]. Our case is an example of implicit voting, since our model will use binary choice data that identifies whether a product is placed in cart or not.

5.3   Model

The model we are proposing is as follows. We are assuming a scenario where all products have RFID tags, and grocery carts are equipped with RFID readers. The carts in the grocery store are equipped with an RFID scanner, which is utilized to collect information on the products that are placed in the customers' cart. In each cart, there is also a small screen where the promotions are displayed. The basic idea of our model is to inform the customer about those products on promotion that are most

likely to be purchased by the customer. The selection of the product advertised is a dynamic decision making process since it is based on the information of the products placed inside the cart while customer is shopping. Collaborative filtering will be used for the identification of the advertised product and Bayesian networks will be used for the application of collaborative filtering.

At the beginning of the shopping process, there are no products in the cart. At this stage, the system can just display those products on promotion that have the highest marginal probabilities. As the customer places products in the cart, the system can display those products that the customer is likely to be interested in purchasing based on items in the cart. The next section describes the data set used for the demonstration of our model.

5.3.1 Netflix Data Set

The proposed model above requires data captured through RFID systems for the different market baskets of the customers. For demonstration of the working mechanism of the proposed model, we used a random selection from the data set available for the Netflix prize competition [Netflix, 2007]. The Netflix prize competition seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on the ratings of the movies they have already seen.

The training data set of the Netflix prize competition constitutes of 17,770 files, one per movie. Each file contains customer ID, the rating given by the customer, and the date of the rating. The ratings are on a scale from 1 to 5, 5 as being the best rating

possible. For the analysis done in this paper, 1,695 movie files from this training data set have been chosen on a random basis. These separate data files are merged into a big data set where the ratings for the movies are sorted based on the customer ID and the date has been dropped out.

The goal of our model is trying to predict the products that the customer may be interested in based on the products that she has placed in the cart. Trying to interpret the customers' behavior suggests the need for implicit voting instead of a detailed 1 to 5 rating scale. Hence, we transformed our data set into a new data set where the ratings 3, 4 and 5 are replaced by 1's as an indicator of the customers' positive preference for the movies. If the customer has rated the movie as 1 or 2 or has not rated the movie at all, then the movie rating is replaced with a zero, which means that it is not in the cart. Here we are assuming that the movies not rated by the customer are movies that are not in the customer's cart.

In a grocery store, there are literally hundreds of thousands of different products. For the problem of finding associations between the products that are in carts, we need to aggragate products. For example, tomato sauce may be sold in different brands, different sizes, different packaging, etc. and all of these need to be aggregated into a single product.[1] The problem of finding a good aggregation can be a difficult one. We need to decide on a number of aggregated products, and a technique to do

---

[1]In the first iteration, we selected 33 movies that had a large number of user ratings (without doing cluster analysis) and used it to learn a Bayes net. However, that was not very effective in predicting the baskets of users in the test set (lift over marginal was about 0.04364)

the aggregation. The optimal number of aggregated products is an empirical question, and an approximate number can be found by experimentation.

After transformation of the data to the desired format, the next step was to select the movies that are going to be used for creating a Bayes net. In order to select the movies from different groupings we used cluster analysis. The FASTCLUS procedure in SAS was used for cluster analysis, where we limited the maximum number of clusters obtained to thirty[2]. As a result, we obtained 30 different clusters and chose one movie from each cluster on a random basis. The final data set used to build the Bayes Net constitutes the movie preferences of 65,535 users for the 30 movies selected. The set of movies selected appears in the Bayes net model shown in Figure 5.1.

Our motivation for learning a Bayes Net is to find the predictive relationships between the movies based on the movies liked or disliked by the customer. WinMine [Heckerman *et al.*, 2000], a tool developed at Microsoft Research, is used to learn a Bayes Net. Using WinMine, the data is divided into a training set and a test set. We performed a 70/30 train/test split and had 45,874 training cases and 19,661 test cases. All of the variables are used as input-output variables (both predicted and used to predict). To set the granularity of the Bayesian network learnt by WinMine, a factor called *kappa* is used, which a number between 0 and 1. As *kappa* approaches 1, the

---

[2] We did not attempt to determine an optimal (or an approximate) number here. We picked thirty for convenience. Since we obtained good results, we did not experiment with other numbers.

model becomes very dense. Since our model is already quite dense, we decreased the value of kappa from its default value of 0.01 to 0.00001. The resulting BN is given in Figure 5.1.

The accuracy of the learned model on the test set is evaluated using the log score Score(x_1, …, x_N) = ($\sum_{i=1}^{N}$ $\log_2 p(x_i|\text{model})$)/$nN$, where n is the number of variables in X, and N is the number of cases in the test set. Our model results in a log score of −0.4169, meaning on average, the log probability that each variable assigns to the given value in the test case, given the values of all other variables , is −0.4169, which translates to a probability of 0.75. Using WinMine we can also compare the difference between the provided model and the marginal model. A positive difference is desired between the provided model and the marginal model, signifying that the model out-performs the marginal model on the test set. The marginal model uses the marginal probabilities of the products in the data set, ignoring the information about the products that are placed in cart. In the same way that a regression model is more accurate than a simple baseline model chosen in the form of a mean dependent value, the "lift over marginal" log score provides information on how well the model fits the data. The lift over marginal log score in our model is 0.1302, which suggests the performance of our model is quite good[3]. If we ignored the products in the cart and used the marginals for prediction, the average probability of the correct prediction is

---

[3] The lift over marginal log score for the Bayesian network model in the WinMine toolkit tutorial is 0.0890. In comparison, our results compares favorably.

0.68 (or log score of −0.5471). Using the products in the cart, the average probability of correct prediction improves to 0.75 (or log scoreof -0.4169) resulting in a lift over marginal log score of (−0.4169)–(−0.5471) = 0.1302. There are many ways of evaluating collaborative filtering recommender systems (Herlocker *et al.*, 2004), and lift over marginal is a good conservative measure of effectiveness for our application.

5.3.1.1   Case Study

In the previous section we have illustrated how a BN can be learned using the WinMine toolkit. Using the probability tables constructed by WinMine we constructed the same Bayes Net in Hugin, a commercial software package. The conditional probability table used for the movie 'Lord of the Rings: The Two Towers' is illustrated in Table 5.1 below.

Table 5.1: The conditional probability table for *Lord of the Rings: The Two Towers*

Lord of the Rings: The Two Towers																
Forrest Gump	0								1							
Titanic	0				1				0				1			
X-Men United	0		1		0		1		0		1		0		1	
Indiana Jones	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	0.88	0.76	0.53	0.21	0.83	0.59	0.37	0.16	0.8	0.52	0.33	0.13	0.7	0.41	0.27	0.1
1	0.12	0.24	0.47	0.79	0.17	0.41	0.63	0.84	0.2	0.48	0.67	0.87	0.3	0.59	0.73	0.9
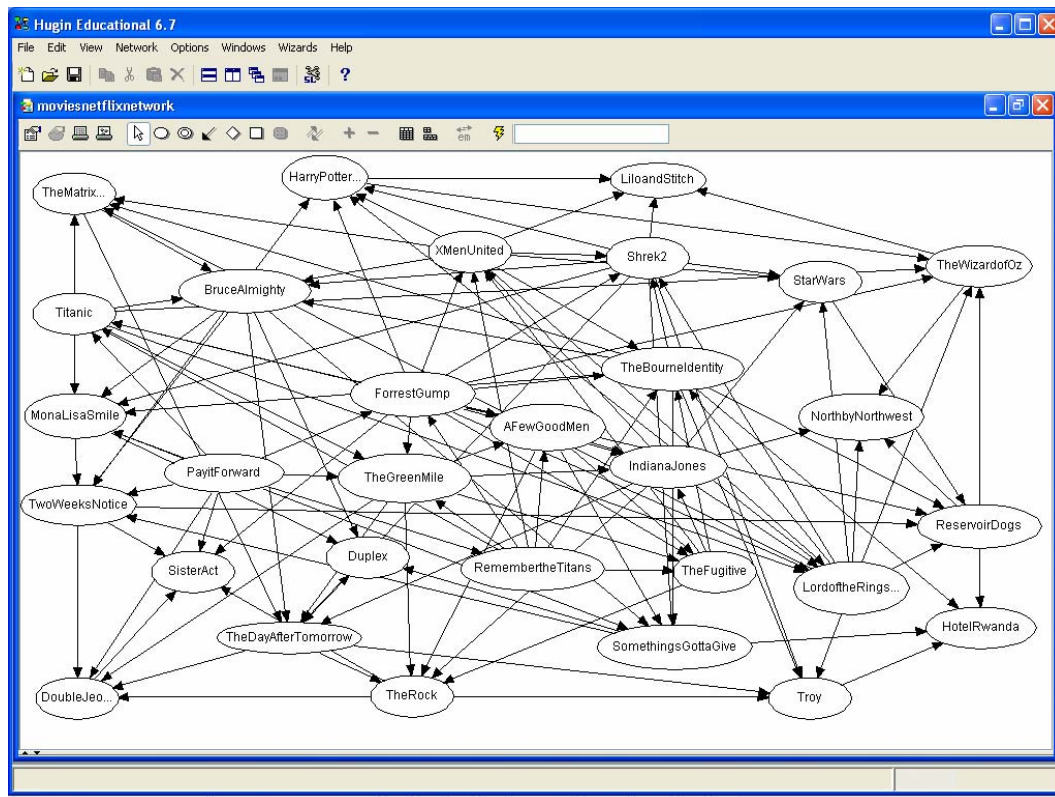
Figure 5.1: A Bayes net for 30 movies from the Netflix prize dataset.

The advantage of using Hugin is that we are able to enter evidence to the BN and update all probabilities accordingly using the 'sum normal' propagation method. In addition to that, the 'max normal' propagation method allows us to find states to the most probable configuration. The state of node with the most probable configuration is given the value of 100. The values for all other states are the relative values of the probability of the most probable configuration in comparison to the most probable configuration.

By using the sum-propagate normal propagation method without entering any evidence, we obtain the marginal probabilities for all the movies in the BN. The results suggest that for the state '1' the movie 'The Green Mile' has the highest

marginal probability 40.57%, and 'Duplex' has the lowest marginal probability 5.23%.

Suppose we want to predict whether a specific customer is going to like the movie *Forrest Gump* or not. Without having any information about the customers' previous movie preferences the marginal probability for the state '1' is 40.43% and the state for the most probable configuration is '0'. Suppose we get the information that the customer rented the movie *A Few Good Men* and liked it. Accordingly, the posterior marginal for *Forrest Gump* increases to 69.75%, the most likely state is still '0'. Next, suppose we get the information that the customer also liked *The Wizard of Oz*. The posterior marginal probability for *Forrest Gump* increases to 90.13% and the most likely state changes to '1'. The results for this case are summarized in Table 5.2 below and the revised Bayes net is given in Figure 5.2 below.
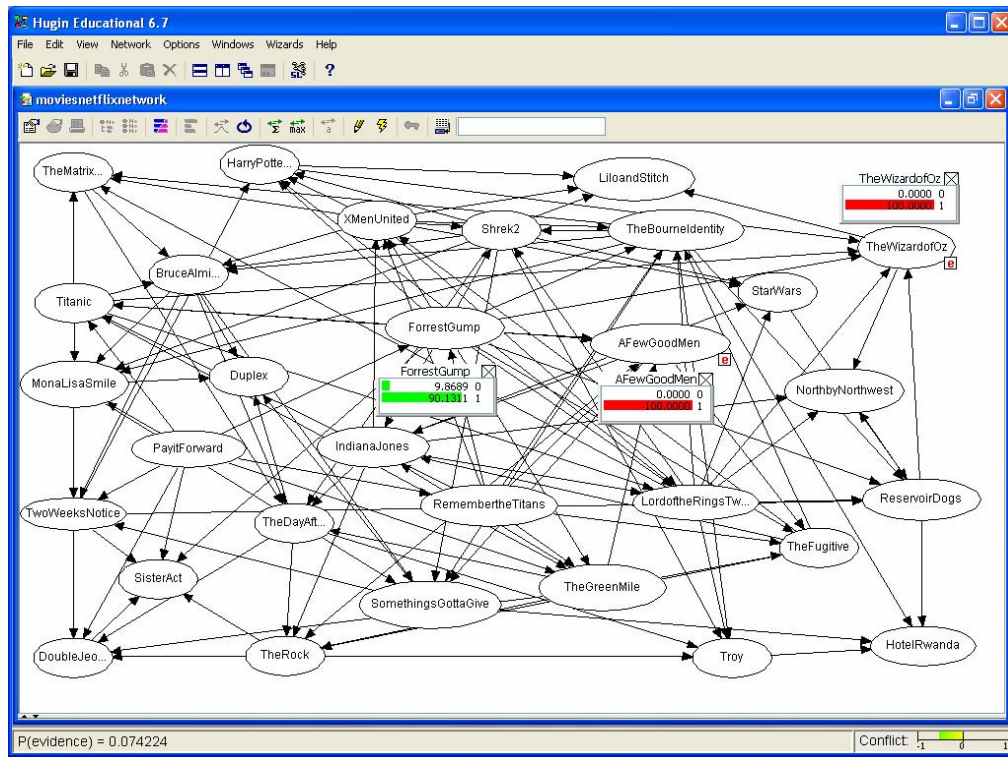
Figure 5.2: The revised Bayes net after entering the evidence

Table 5.2: Posterior probabilities and most likely state for *Forrest Gump*

Information & Rating	Marginal	Most likely state
Prior	40.43%	0
A Few Good Men = 1	69.75%	0
Wizard of Oz =1	90.13%	1

As our second case, consider a scenario where we need to choose between the two movies *Mona Lisa Smile* and *Lord of the Rings*: *The Two Towers* to recommend to the customer. The initial most likely state is '0' for both movies. Based on their marginal probabilities, which are given in Table 5.3 below, *Lord of the Rings: The*

Two Towers should be chosen for recommendation, since it has a much higher marginal probability for the state '1'.

Table 5.3: Posterior probabilities and most likely states for *Mona Lisa Smile* and

Lord of the Rings: The Two Towers

Information & Rating	Mona Lisa Smile		Lord of the Rings: The Two Towers	
	Marginal	Most likely state	Marginal	Most likely state
Prior	19.28%	0	33.85%	0
Pay It Forward = 1	36.86%	0	48.08%	0
Something's Gotta Give = 1	63.05%	0	61.37%	0
Two Weeks Notice = 1	72.22%	1	63.23%	0
Titanic = 1	76.00%	1	67.38%	1

Suppose we receive information about movie preferences of the customer to whom we are going to make the recommendation. Learning that the customer liked *Pay It Forward*, *Something's Gotta Give*, *Two Weeks Notice* and *Titanic* with the particular order given, changes the posterior marginal probabilities. Until we obtain the information that the customer liked *Something's Gotta Give*, the marginal probabilities indicate that *Lord of the Rings* should be chosen for recommendation. After subsequent observations, *Mona Lisa Smile* takes the lead for recommendation. At the point where we learn that the customer liked *Titanic*, the most likely state for

both of the movies becomes '1' where for *Lord of the Rings* it is '0' still. After we get

the information that the customer also liked the movie *Titanic* the most likely state for

both of the movies becomes'1'. The details of posterior marginal probabilities and the

most likely states are given in Table 5.3 above. The revised Bayes Net is given in

Figure 5.3 below.
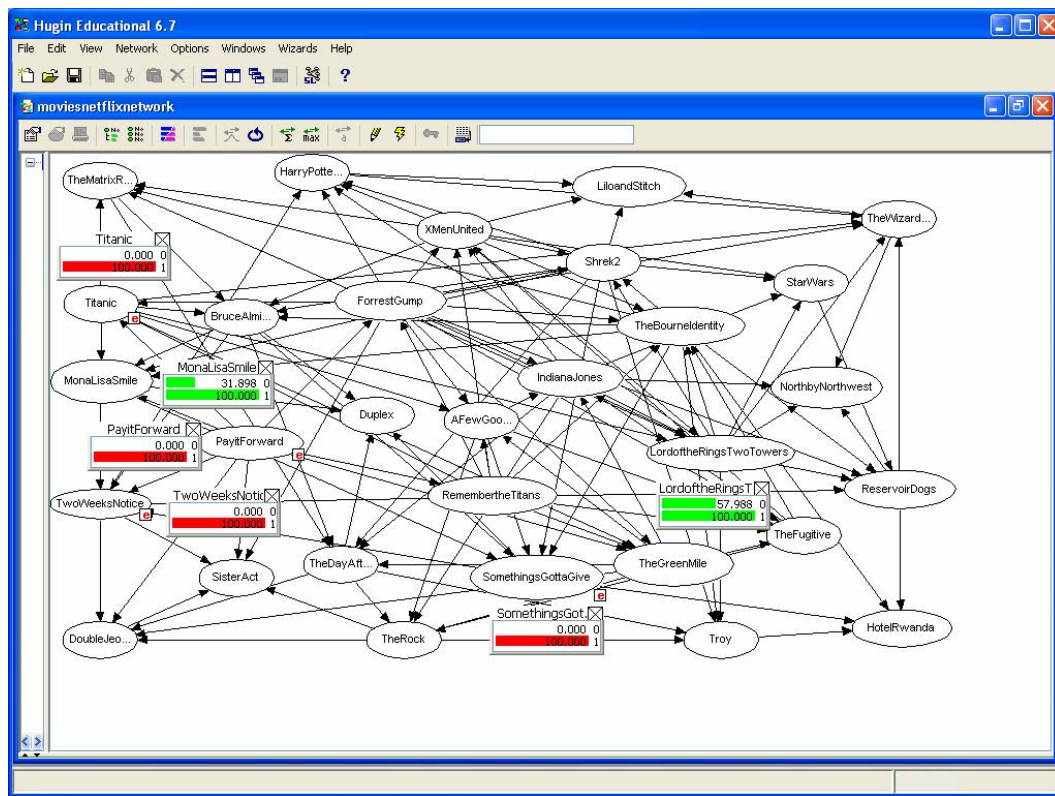
Figure 5.3: The revised Bayes net after entering the evidence in the second case

5.3.2   Grocery Data Set

In the first part of this research we illustrated our application using the publicly

available Netflix price competition data set. While it is not quite the same, we were

able to convert the Netflix dataset with movies as the products instead of grocery

items. As described in section 5.3.1, the Netflix data set constitutes of the ratings given by the users. Ratings data sets can be quite informative for the researchers since they provide information about the preferences of the subjects of the data pool. On the other hand, the collection of the ratings data sets is mostly based on the voluntary participation of the users, which causes the actual data pool to be veiled when there is a lack of ratings. This fact about the ratings data set forced us to make the assumption in our analysis, that the movies not rated by the customer were movies that were not in the customer's cart.

As the next step of this research, we are going to construct our proposed model using real grocery data which enables us to eliminate the assumption made before, hence obtaining a more accurate representation of the application we are proposing.

The data set we are going to use is a retail market basket data set, supplied by an anonymous Belgian retail supermarket store [Brijs *et al*, 1999]. The data has been collected over three non-consecutive periods, resulting in approximately 5 months of data. Each line in the data set corresponds to a shopping basket. The distinct items in the data set are encoded by numbers[4], but the information to which products those numbers correspond is not provided in the data set. We also do not know how many of the same product is present in a single shopping basket.

---

[4] Although most of the products are identified by a unique barcode, some article numbers in the data set represent a group of products rather than an individual product item [Brijs. *et al*, 1999].

In total, the data set contains the product information of 51,273 different shopping baskets which has been purchased by 5,133 different customers. The number of distinct products included in the data set is 14,472. The average number of distinct items per visit is 13 but in most of the shopping baskets there are about 7 to 11 distinct items.

Remember for our analysis, we are only interested to know whether a product is present in the shopping basket or not. The quantity of the products purchased does not matter for us. In order to be able to do our analysis, we either had to create a huge matrix where rows correspond to 51,273 different shopping baskets and columns correspond to the different 14,472 products or we had to transform the data into a form where there are only two columns, the basket ID and the product ID. In order to achieve efficiency in our analysis we selected the second choice. After transformation, every single row in the data set gives information about one shopping basket and one product that is placed inside. Thus, the same basket can be seen in as many rows as the number of products in it.

Using the entire data set to construct the BN is not possible since the huge number of different products correspond to 14,472 variables in the BN. To learn a BN of that size, a much higher quantity of market basket data is necessary than it is available in this data set.

Remember in the first part of this research we used cluster analysis as the aggregation technique. The basic idea of our approach was that people tend to watch different kinds of movies depending on various reasons (like the mood they are in).

For that reason we chose one movie from each cluster on a random basis and created our BN using those movies. The results we got using this approach proved to be successful. We got a lift over marginal of 0.1302 which corresponds to about 7% improvement in our predictions. However, grocery data possess different characteristics compared to the Netflix data set which suggests our approach for analyzing the data needs to be updated.

The way customers tend to select the products in a single shopping is different than the way they select movies. Depending on the purpose of a single grocery shopping the products selected may tend to belong to the same grouping. In this context, aggregation techniques might help for a better analysis, nevertheless the information we have in the grocery data set is much more limited compared to the Netflix data set. In the Netflix case we were fully aware of the movies included in the data set. On the contrary, by the grocery data set we do not have any information about the products that are present in the data set. The grocery data set includes the information of 51,272 market baskets which include 14,472 distinct products. But the number of the different customers these 51,272 market baskets refers to is only 5,133 and the majority of products are only present in a few baskets whereas five products with the highest purchase frequencies amount to about 16% of the whole data set. On the other hand, for the Netflix data set, the number of different users was 65,535, providing a much larger scale of information about users' preferences. Considering all of these, we conclude that cluster analysis is not the right method and a different

approach needs to be used to select the products that are going to be used to create the BN.

The purpose of this study is to recommend the products to the customer in which (s)he will be interested and which is likely to be purchased by the customer. For that reason, we use Bayes nets to learn the predictive relationships between the products. However, the information gathered from a BN is closely related to the structure of the data set used to learn that BN. For grocery data sets; it is almost always the case that some products like milk, egg and bread, are purchased on a regular basis by the customers. As a result, they occupy the top of the list as the most purchased products in the data set, leaving a wide gap behind for the followers. A BN created from such a data set, will naturally show high dependencies between those most purchased products and many others in the data set. This is an expected result, for that those products constitute a big part of the whole data set. Referring this as a reference on the other hand, may result in making recommendations of milk or eggs for every possible kind of product placed in the shopping cart. It should be noted that, being purchased at the same shopping trip does not always necessarily show the tendency for those products to be purchased together by the customers. It may just be the result of high and uneven (compared to other products) purchase frequencies of those products that are purchased on everyday basis.

When we look to our grocery data set we see the picture described above. There is a huge gap between the frequencies of the top five most purchased products and the rest of the data set. About 55% of the 14,472 distinct products in the data set are

present in only 10 baskets or less. In contrast, this number is 82,040 for the top five products in total[5]. Anticipating these are the products that are bought on a daily basis (like milk, egg etc) we are going to exclude them from the selection process of the products to create the BN.

With this application proposed, we aim to make recommendations based on the products placed in cart. Hence, in order to achieve an accurate demonstration of a single grocery shopping the products used in the BN should be representative for a market basket, meaning that they are likely to be purchased on the same shopping trip. For it is necessary to update our approach because of the different characteristics of the grocery data set, we will use the conditional probability tables obtained through the BNs we create to see the associations between the products, hence to select the products to be used in the final Bayes net we create.

The Bayes nets inform us about the conditional independencies and the conditional probabilities of the variables. Thus, for each variable used in a BN, we may obtain the conditional probability table associated with that variable. Our basic idea is to use this information contained in the BNs as a selection basis to identify the products that have high associations with each other; the ones that are likely to be purchased on the same shopping trip.

We proceed as follows: The first step of the heuristic is to select a certain number of products that are going to be used to create the BN. In this case, excluding the top

---

[5] This number is the sum of each of these products' count for the baskets they are in.

five, we selected the 25 products that have the highest purchase frequencies, so we have enough data available to build the BN. This way we avoid resulting in poor recommendations as a consequence of the high and uneven frequencies of the daily consumed products. The BN created with those 25 products is given in Figure 5.4 below.
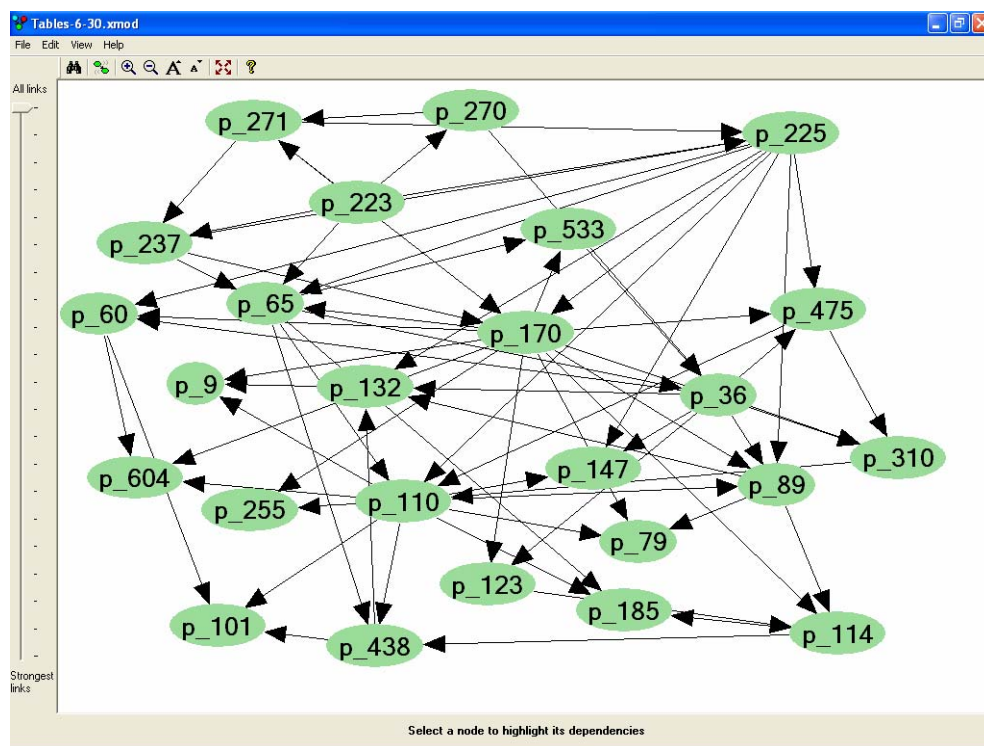
Figure 5.4. The BN created using 25 products with high purchase frequencies

The main purpose to create this BN is to obtain the information about the dependencies between the products and to learn the conditional probability tables. Using conditional probability tables we can observe the change on the purchase probabilities of a child node depending whether the parent product(s) is present in

cart or not. This change is high for some products, indicating that they are highly associated with their parents and low for some, as a result of weak associations.

Consider the conditional probabilities of the product *p_271* given in Table 5.4. The purchase probability of the *p_271* increases from 0.037 to 0.365 when *p_*270 is included in the basket. The purchase probability of *p_*271 increases even to 0.418 when the product *p_2238* is placed in the basket. These high changes in the purchase probabilities of *p_271* suggest that this is an example of a product which is highly associated with its parents.

In contrast, consider the case illustrated in Table 5.5. The conditional probabilities of product *p_533* indicate that the purchase probability of *p_533* drop when the parent products, *p_170* and *p_65*, are included in the basket. This is a sign of weak associations of the child product with its parents indicating that it is not very likely that they are going to be included in the same shopping basket.

Table 5.4: The conditional probability table for *p_271* given its parents

Probability table for *p_271*			
p_2238	*p_*270	(*p_271* = 0)	(*p_271* = 1)
0	0	0.963	0.037
0	1	0.635	0.365
1	0	0.882	0.118
1	1	0.582	0.418

Table 5.5: The conditional probability table for *p_533* given its parents

Probability table for p_533			
p_170	*p_65*	*(p_533 = 0)*	*(p_533 = 1)*
0	0	0.956	0.044
0	1	0.981	0.019
1	0	0.988	0.012
1	1	0.986	0.014

A way to measure the degree of change is to apply the distance formula

$$\sum_i (p_i - q_i)^2 / \binom{n}{2}$$ for every possible pairs of the conditional probabilities and then

average them. In this formula, *p* and *q* stand for the conditional probabilities of the

child node for the different states of its parents *i* stands for the different states of the

child node and *n* stands for the number of states of the set of parent nodes. A high

average distance is desired as an indication of the high association of the child node

with its parents and can be used as a reference to select the products used in the final

BN created. The average distances of *p_271* and *p_533* are calculated as 0.068711

and 0.0004 respectively. With *p_271* having a higher average distance than *p_533*,

our prior conclusions are verified.

However, the selection of the products shouldn't be based on average distances

solely since there may be products which are poorly associated with its parents but

highly associated with its children. As an example for that, consider the case illustrated in Figure 5.5. Product *p_270* has *p_2238* as its parent and *p_271* as its child. Looking at the conditional probability table illustrated in Table 5.6, the average distance of *p_270* from its parent is calculated as 0.0013482, which suggests that *p_270* should not be selected to the final BN to create. However, from our prior findings we know that having *p_270* in the cart has a huge impact on the purchase probability of *p_271*[6]. Not selecting a product like that would just deteriorate our findings.
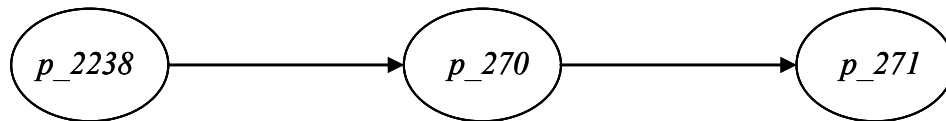
Figure 5.5. The BN created using 25 products with high purchase frequencies

Table 5.6: The conditional probability table for *p_270* given its parent

Probability table for p_270		
p_2238	*p_270 = 0*	*p_270 = 1*
0	0.957965	0.0420349
1	0.841851	0.158149

Ideally, we should find the distance in conditional probabilities of a parent node with its children, sum it with the average distance of the same node with its parents and use the result obtained to select the products for the BN. Nevertheless, although

[6] Look at Table 5.4

the capture of this information is possible, it requires time consuming calculations. For that reason, the selection process will be materialized following the heuristic described below.

Notice that, the average distance in the conditional probabilities obtained through BNs shows the associations of the child node with its parents. This is another way of saying that it shows the association of a parent node with its child, jointly with the child's other parents. Hence, the average distance in conditional probabilities can also be used as a reference for the level of association of a node with its child. Accordingly, as the first step of the heuristic, we will calculate the average distance of each variable's conditional probabilities. Let d_j denote this average distance, where j denotes variable considered. Using d_j, a score called S_j will be calculated which is the sum of the distances of the variable of interest and its children. Thus, $S_j = d_j + \sum_i d_{ij} / C_j$, where i denotes the child node(s) of the variable j and C denotes the number of j's children. S_j will be used as a reference to select the products that are going to be used in the final BN. Calculation of S_j for the product p_36 is illustrated in Figure 5.6 below.
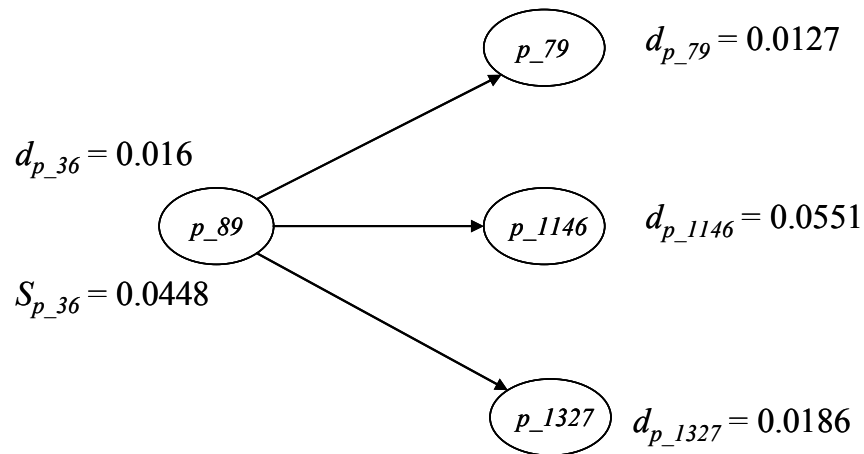
Figure 5.6. Calculation of S_j for the product p_36

The heuristic is as follows: The variable with the lowest S_j is the variable that has the lowest level of associations with the other variables. Hence, that variable will be excluded from the analysis. With the variables remaining, a new BN will be created and from the conditional probability tables obtained through this new BN new S_j scores will be calculated. These steps, (exclusion of the variable with the lowest S_j score and creation of the new BN) will be repeated until we end up having as many variables remaining as the number of variables that we want to use in the final BN.

Using this heuristic we will be able to find the products that have high associations with each other, hence the products that are likely to be purchased on the same shopping trip. In addition to that, with the help of this heuristic we are able to detect the association between the products in the data set, which are not present to a very high degree in the database and may be overlooked because of that in the analysis of the aggregate data. On the other hand, the current problems involved with this heuristic are mostly because of some technical difficulties. We do not have a system to automate the procedure. We obtain the cpt's from the BNs created in

WinMine. However, WinMine does not allow copying the cpt's, so after printing the tables and recreating them in Excel we find the average distances. Because of this problem involved, we chose the top 10 products with the highest S_j scores instead of discarding the variables with the lowest S_j scores one by one. With the BN obtained this way, we got a lift over marginal of 0.229405 which corresponds to 13.4829% improvement between the provided model and the marginal model. The BN created is given in Figure 5.7 below.
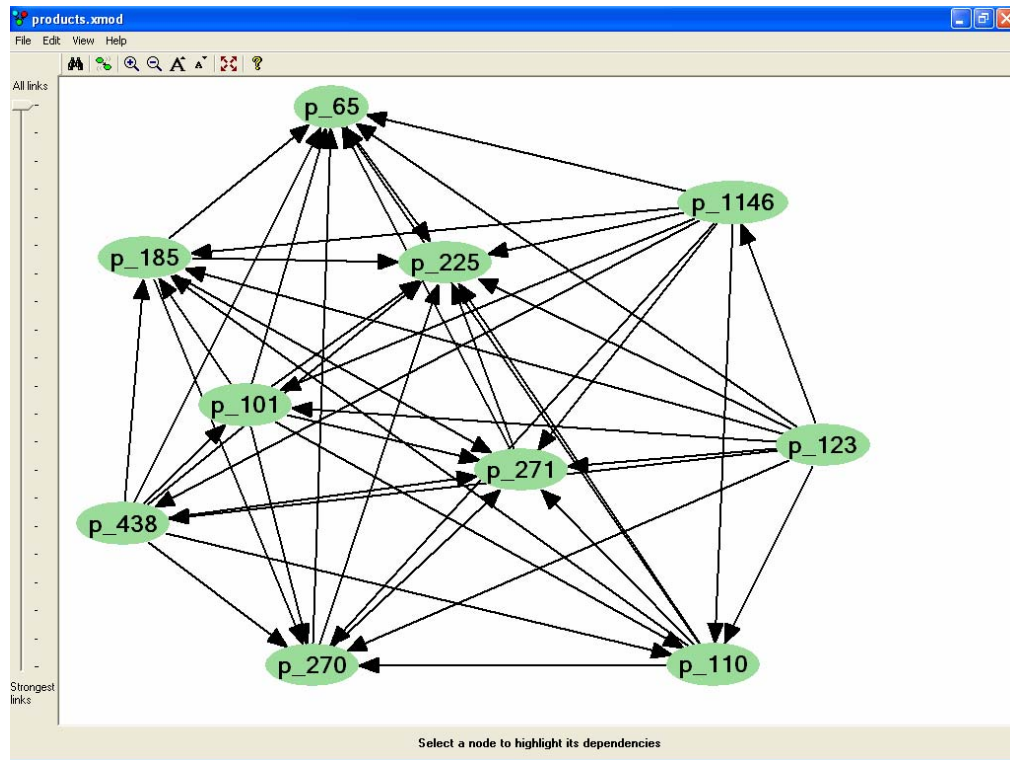
Figure 5.7. The BN created using 10 products with highest S_j

5.3.2.1   Case Study

In the previous section we illustrated the heuristic we used to create the BN of interest. Using that heuristic we created a BN with 10 products using WinMine toolkit. As the next step, using the probability tables constructed by WinMine we are going to construct the same BN in Hugin. That way we are going to be able to obtain marginal probabilities of the products. In addition, using the methods like 'sum normal' propagation and 'max normal' propagation in Hugin, we can enter evidence, update all probabilities accordingly and find states to the most probable configuration.

The conditional probability table used for the product *p_101* is illustrated in Table 5.7 below. Notice that there are two initial states for each product, '0' and '1', representing that the product is 'not in cart' or 'in cart', respectively.

Table 5.7 The conditional probability table of *p_101*

				p_101				
p_123			0				1	
p_438		0		1		0		1
p_1146	0	1	0	1	0	1	0	1
0	0.860	0.975	0.985	0.888	0.968	0.857	0.956	0.5
1	0.140	0.025	0.015	0.112	0.032	0.143	0.044	0.5

By using the sum normal and max normal propagation methods without entering any evidence, we obtain the marginal probabilities for all the products and the states belonging to the most probable configuration. Accordingly, the product *p_65* has the highest marginal probability 23.16% for the state '1', and the product *p_123* has the

lowest marginal probability 6.868% for the state '1'. Other than the product *p_65*, the most likely state for all the remaining products is '0'. The marginal probabilities and the most likely state of each product can be seen in Table 5.8 given below.

Table 5.8. Marginal probabilities and most likely states of the products in the BN

	Marginal for '0'	*Marginal for '1'*	*Most likely state*
p_65	76.840 %	23.160 %	1
p_101	88.901%	11.099 %	0
p_110	86.3 %	13.7 %	0
p_123	93.132 %	6.868 %	0
p_185	92.151 %	7.849 %	0
p_225	81.912 %	18.088 %	0
p_270	90.352 %	9.648 %	0
p_271	88.478 %	11.522 %	0
p_438	90.158 %	9.842 %	0
p_1146	91.351 %	8.649 %	0

Suppose we want to predict whether a specific customer will be interested in product *p_271* or not. Without any other information the marginal probability for the state '1'is 11.522 % and the state belonging to the most probable configuration is '0'. According to these results it is not very much likely that the customer is going to like the product *p_271*. However, suppose through our RFID scanner we get the information that the customer placed product *p_270* in his cart. Accordingly, the posterior marginal of product *p_271* for the state '1' increases to 37.635 %. The most likely state remains '0' still. Next, suppose we get the information that the customer also replaces the product *p_123* in his cart. The posterior marginal probability for the

product *p_271* increases to 55.080 % and the most likely state becomes '1'. The results for this case are summarized in Table 5.9 below.

Table 5.9. Posterior probabilities and most likely state for product *p_271*

Information	Marginal	Most likely state
Prior	11.522 %	0
p_270 = 1	37.635 %	0
p_123 = 1	55.080 %	1

As the second case, consider a scenario where two promotions are going on inside a grocery store, a promotion for the product *p_225* and a promotion for the product *p_65*. The task of the manager is to decide to which customer to advertise which one of these two products. Based on the marginal probabilities, which are given in Table 5.10 below, the product *p_65* should be chosen since it has a higher marginal probability. But being able to observe what the customers are placing in their carts, we may change our recommendation. Suppose we observe a customer, which placed the products *p_110* and *p_123* in his basket. Accordingly, we decide to recommend *p_225* to that particular customer, since after observing what is placed inside the cart the posterior marginal for *p_225* increases to 36.443%, where for *p_65* it is 23.189%. The next customer we observe has three products placed inside her cart, *p_ 101*, *p_271*, *p_1146*. With these products placed inside the cart the posterior marginal probability for *p_65* increases to 56.170 %. Hence, *p_65* will be recommended to the customer. The details are given in Table 5.10 below.

Table 5.10. The posterior probabilities and most likely states for *p_225* and *p_65*

Information	p_225		p_65	
	Marginal	*Most likely state*	*Marginal*	*Most likely state*
Prior	18.088 %	0	23.160 %	1
Cart 1(p_110 = 1, p_123 = 1)	36.443 %	0	23.189 %	0
Cart 2 (p_101= 1, p_271 = 1, p_1146 =1)	45.887 %	0	56.170 %	1

5.4   Summary and Conclusions

RFID is a technology that has a huge potential for commercial applications. This huge potential of RFID technology is present and known by the researchers. However, we are in need for new ideas which uncover this potential and transform it into managerially useful information. By using RFID and collaborative filtering for targeted advertising in grocery stores, the application illustrated in this research, promises to do so. With this application we are able to transfer the methods of e-commerce to actual retail stores, which allow us to influence the customer's decision making process while shopping. With this application, through real time data collection with RFID we will be able to increase the success of the promotions and furthermore as a future research topic we can get insight about the operational problems, such as the optimal placing of products inside a store.

In this research, the application proposed is illustrated first using the data for the Netflix prize competition where we converted the Netflix data set with movies as the products instead of grocery items. While the Netflix data set is very rich in many aspects (like the number of users contained, the number of movies, information about the movies *etc.*), the fact that it is based on the ratings forced us to make the assumption that the movies not rated by the customer were movies that were not in the customer's cart. As the second stage of the research, real grocery data is used which helped us to eliminate this assumption made before. The limitations involved in the grocery data set available, (not having information about the products contained in the data set, not having enough market basket data available compared to the number of distinct items in the data set *etc.*) gave us the incentive to develop a new heuristic. Using the methods of this new heuristic we were able to expose the information contained in a limited data set to find the products that have high associations with each other, hence the products that are likely to be purchased on the same shopping trip. In addition to that, this new heuristic may detect the association between the products in the data set, which are not present to a very high degree in the database and may be overlooked because of that in the analysis of the aggregate data.

Many grocery stores have data on users using loyalty cards. As a future research topic, the longitudinal information about these users can be used to improve the effectiveness of our system.

6 SUMMARY and CONCLUSIONS

6.1   Summary and Conclusions

Two essays are presented in this research. In the first essay "On solving stochastic PERT Networks" first we review the methods used to solve stochastic PERT networks. The current methods in the literature fail to recognize the true distribution of the maximum of two independent distributions and therefore make false assumptions about activity distributions, like the maximum of two normal distributions are again normally distributed. Depending on the value of parameters this assumption can lead to large errors for the completion time of the activities which leads to inaccurate estimates for the project completion time.

Motivated by this problem in the literature, in Chapter 2 we provided a new method which aims to approximate the true distribution of the project completion time by eliminating the false assumptions for the distribution of the maximum of two Gaussians. In this method, a PERT network is first transformed into a MoG Bayesian network and then Lauritzen-Jensen algorithm is used to make inferences in the resulting MoG Bayesian network. In Chapter 1, first a PERT network is transformed into a PERT Bayesian network [Jenzarli, 1995] which allows us to depict the dependencies between the activities. Following that in subsection 2.3.1 we demonstrated how non-Gaussian distributions can be approximated using MoG's. In section 2.3.2 we investigated the problem of finding the distribution of maximum of

two Gaussian distributions and shown that using MoG's an accurate estimation for the maximum of two distributions can be obtained.

In MoG Bayesian networks the discrete variables can not have continuous parents. Consequently as we convert PERT networks into MoG Bayesian networks arc reversals become necessary. This fact motivated us to work on the arc reversal theory. As a result, we described arc reversals between every possible pairs of variables. We also described a new kind of distribution called partially deterministic that can arise in the process of arc reversals in hybrid Bayesian networks.

Because of the restrictive nature of MoG Bayesian networks transforming a PERT network to a MoG Bayes net is too complex for practical use. As an alternative to MoG's, in Chapter 3 we explored solving stochastic PERT networks using mixtures of truncated exponentials. MTE potentials can be used for inference in hybrid Bayesian networks that do not fit the restrictive assumptions of the conditional linear Gaussian (CLG) model. We demonstrated the easy applicability of MTE potentials by finding the marginal probability distribution of a PERT example using MTE's. This calculation process involves the conversion of the PERT network into a PERT Bayes net, transformation of the PERT Bayes net into a MTE network and finally propagation of the MTE potentials using the Shenoy-Shafer architecture. Finding the distribution of the maximum of two distributions using MTE's are first described here as an operation necessary to propagate in MTE PERT networks.

Finally, we evaluated our method by compared the mean and variance of the marginal distribution of the project completion time with the exact analytic results

using Clark's method [1961] and by comparing the shape of our distribution with the actual distribution calculated by brute force using order statistics. In both of these cases our method proved to be successful.

The second essay "Using RFIDs for Operations Management" first investigates the potential of RFID in terms of operations management. In this research we state that the real potential of RFID systems is hidden in the massive amounts of data that is captures through RFID systems. In this context, we argued that researchers should be looking for ways to transform this data into managerially useful information.

The application we are proposing in this research is an attempt to uncover the real potential of RFID systems and consequently transfer the methods of e-commerce to actual retail stores. In our model RFID systems are used to collect real time information of the products placed inside a cart by customer during the shopping. With this information handy, collaborative filtering is used for the identification of the product to be advertised to the customer and Bayesian networks is used for the application of collaborative filtering. In this way we are able to influence the customer's decision making process while shopping and to inform the customer about the promotions in store in which the customer is likely to be interested.

The application proposed is illustrated using two different data sets, the data set available for the Netflix prize competition where we converted the data set with movies as the products instead of grocery items and a grocery data set. Netflix data set is based on the ratings of the customers and that forced us to make the assumption that the movies not rated by the customer are movies that were not in customer's cart.

Using grocery data as the second stage of this research we were able to eliminate the assumption made before and hence obtain a more accurate representation of the reality. The limited information contained in the grocery data set motivated us to develop a new heuristic to find the products that have high associations with each other, hence the products that are likely to be purchased on the same shopping trip.

We used WinMine to learn the Bayes net and evaluated our models using the lift over marginal log score. The lift over marginal obtained with our data sets are 0.1302 and 0.229405 for the Netflix data set and grocery data set respectively, meaning that the average probability of correct prediction improved about 7% with the model builtusing the Netflix data set and about 13.4829% with the model built using grocery data set. Thus, in both of these cases our models proved to be successful.

6.2   Directions for Future Research

With this research it is shown how the false assumptions involved in the computation of PERT networks can be eliminated and thus an accurate estimation of the project completion time can be obtained. As a future research topic a PERT network can be solved both using the methods described here and the current methods used in the literature, representing how large the errors for the project completion time estimates could be if the correct methods for the estimation of the project completion time is not in use. Additionally, we can work on a method to avoid arc reversals in MoG Bayesian networks, thus making our approach more friendly for practical use.

With this research, we demonstrated arc reversals between every possible pair of variable. The arc reversal theory facilitates the task of approximating general Bayesian networks with mixtures of Bayesian networks arc reversals but besides that it is potentially useful to solve hybrid influence diagrams.

Many stores have data on users using loyalty cards. As a future research topic the longitudinal information about these users can be used to improve the effectiveness of our system. Additionally, our model can be altered by considering a grocery store which is equipped with 'smart shelves'. The smart shelves can determine the position of each product that is placed on them and can recognize when products placed on it or removed from it [Decker, 2003]. Using this setting we can obtain information about the optimal placing of products inside a store.

APPENDIX

A Converting the PERT Bayes Net to a MoG Bayes Net Using Arc Reversals

Notice that the example of the PERT Bayes Net, given in Figure 1.2, is not a MoG BN. The variables C_{23} and F are conditionally deterministic variables that are not linear, and D_5 has a non-Gaussian distribution. Using the techniques described in sections 2.3.1 and 2.3.2, these variables need to be approximated by MoG distributions. For that reason, as illustrated in Figure A1 below, three discrete variables (B_5, B_{45}, and B_{23}) with the appropriate distributions are added to the current PERT Bayes Net.
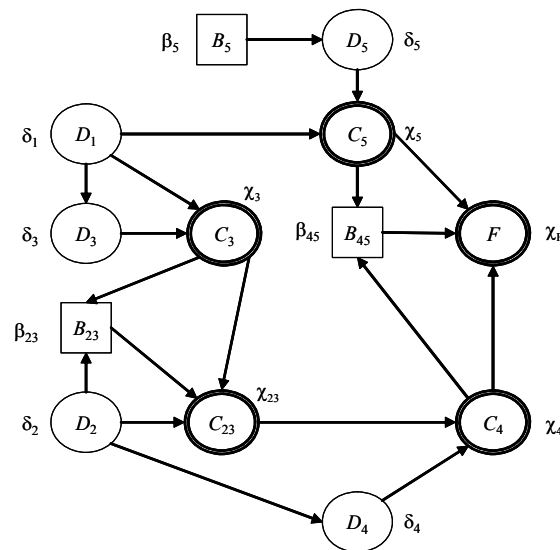

Figure A1: Bayes Net model

For it is easier to show the calculations necessary, the notation of mixed potentials will be used in this section. In hybrid Bayesian networks, we encode conditionals by

mixed potentials that have two parts, a mass part and a density/equation part. The conditional associated with a discrete variable will have a mixed potential with vacuous density part denoted by ι, and the conditional associated with a continuous variable will have a mixed potential that has a vacuous mass part denoted by 1. Mixed potentials can be combined, divided, and marginalized. If $\alpha = (\alpha_1, \alpha_2)$, and $\beta = (\beta_1, \beta_2)$ are two mixed potentials with mass parts α_1 and β_1, respectively, and density parts α_2 and β_2, respectively, their combination is $\alpha \otimes \beta = (\alpha_1 \otimes \beta_1, \alpha_2 \otimes \beta_2)$, and their division $\alpha \% \beta = (\alpha_1 \% \beta_1, \alpha_2 \% \beta_2)$ (assuming that the domain of α_i contains the domain of β_i for $i = 1, 2$). Marginalization of a mixed potential is more complex.

If $\alpha = (\alpha_1, \alpha_2)$ is a mixed potential for a such that α_1 is a potential for a_1, and α_2 is a potential for a_2, $a_1 \cup a_2 = a$, and $X \in a$, then

$$\alpha^{-X} = (\alpha_1^{-X}, \alpha_2) \qquad \text{if } X \in a_1 \text{ and } X \notin a_2;$$

$$= (\alpha_1, \alpha_2^{-X}) \qquad \text{if } X \notin a_1 \text{ and } X \in a_2;$$

$$= ((\alpha_1 \otimes \alpha_2)^{-X}, \iota) \qquad \text{if } X \in a_1, X \in a_2, \text{ and } a \backslash \{X\} \text{ are all discrete (or mixed);}$$

$$= (1, (\alpha_1 \otimes \alpha_2)^{-X}) \qquad \text{if } X \in a_1, X \in a_2, \text{ and } a \backslash \{X\} \text{ are all continuous (or mixed);}$$

If $a \backslash \{X\}$ is a mixed set of variables, then $((\alpha_1 \otimes \alpha_2)^{-X}, \iota)$ and $(1, (\alpha_1 \otimes \alpha_2)^{-X})$ are equivalent and there is no contradiction in the definition of marginalization. In the next section, we will provide several concrete examples of operations with mixed potentials.

The potentials of the variables in the PERT Bayes Net given above are as follows:

Let δ_1 denote the mixed potential at D_1. Thus, $\delta_1(d_1) = (1, \varphi_{3,1}(d_1))$

Let δ_2 denote the mixed potential at D_2. Thus, $\delta_2(d_2) = (1, \varphi_{14,3}(d_2))$

Let δ_3 denote the mixed potential at D_3. Thus, $\delta_3(d_1, d_3) = (1, \varphi_{5+2d_1,2}(d_3))$

Let δ_4 denote the mixed potential at D_4. Thus, $\delta_4(d_2, d_4) = (1, \varphi_{1+d_2,4}(d_4))$

Let β_5 denote the mixed potential at B_5. The details of the potential are as follows:

$\beta_5(b_{51}) = (0.051, \iota)$

$\beta_5(b_{52}) = (0.135, \iota)$

$\beta_5(b_{53}) = (0.261, \iota)$

$\beta_5(b_{54}) = (0.341, \iota)$

$\beta_5(b_{55}) = (0.212, \iota)$

Let δ_5 denote the mixed potential at D_5. The details of the potential are as follows:

$\delta_5(b_{51}, d_5) = (1, \varphi_{0.032,\,0.0112}(d_5))$

$\delta_5(b_{52}, d_5) = (1, \varphi_{0.143,\,0.0482}(d_5))$

$\delta_5(b_{53}, d_5) = (1, \varphi_{0.415,\,0.1382}(d_5))$

$\delta_5(b_{54}, d_5) = (1, \varphi_{1.014,\,0.3382}(d_5))$

$\delta_5(b_{55}, d_5) = (1, \varphi_{2.300,\,0.7672}(d_5))$

Let β_{23} denote the mixed potential at B_{23}.

$\beta_{23}(b_2, c_3, d_2) = (1, \iota)$ if $c_3 - d_2 \leq 0$

$= (0, \iota)$ if $c_3 - d_2 > 0$

$$\beta_{23}(b_3, c_3, d_2) = (0, \iota) \qquad \text{if } c_3 - d_2 \leq 0$$

$$= (1, \iota) \qquad \text{if } c_3 - d_2 > 0$$

Let β_{45} denote the mixed potential at B_{45}.

$$\beta_{45}(b_4, c_4, c_5) = (1, \iota) \qquad \text{if } c_5 - c_4 \leq 0$$

$$= (0, \iota) \qquad \text{if } c_5 - c_4 > 0$$

$$\beta_{45}(b_5, c_4, c_5) = (0, \iota) \qquad \text{if } c_5 - c_4 \leq 0$$

$$= (1, \iota) \qquad \text{if } c_5 - c_4 > 0$$

The variables C_3, C_4, C_5, C_{23} and F are conditionally deterministic variables, so there are no conditional density functions for them. Let χ denote the equations for these variables, so the equations of these variables are as follows:

$$\chi_3 : C_3 = D_1 + D_3$$

$$\chi_4 : C_4 = C_{23} + D_4$$

$$\chi_5 : C_5 = D_1 + D_5$$

$$\chi_{23} : C_{23} = D_2 \quad \text{if } B_{23} = b_2$$

$$C_{23} = C_3 \quad \text{if } B_{23} = b_3$$

$$\chi_F : F = C_4 \quad \text{if } B_{45} = b_4$$

$$= C_5 \quad \text{if } B_{45} = b_5$$

In MoG Bayes nets, discrete nodes cannot have continuous parents. Looking to our PERT Bayes Net example we see that B_{23} has two continuous parents C_3 and D_2, and B_{45} has two continuous parents C_4 and C_5. In order to address this situation we are

going to use arc reversals. The details of each arc reversals will be given on the

following page

Reversal of arc (D_2, B_{23})

Figure A3: After reversing (D_2, B_{23})

$(\delta_2 \otimes \beta_{23})(d_2, b_2, c_3) = (1, \varphi_{14,3}(d_2))$ if $c_3 - d_2 \leq 0$

$(d_2, b_2, c_3) = (0, \varphi_{14,3}(d_2))$ if $c_3 - d_2 > 0$

$(\delta_2 \otimes \beta_{23})(d_2, b_3, c_3) = (0, \varphi_{14,3}(d_2))$ if $c_3 - d_2 \leq 0$

$(d_2, b_3, c_3) = (1, \varphi_{14,3}(d_2))$ if $c_3 - d_2 > 0$
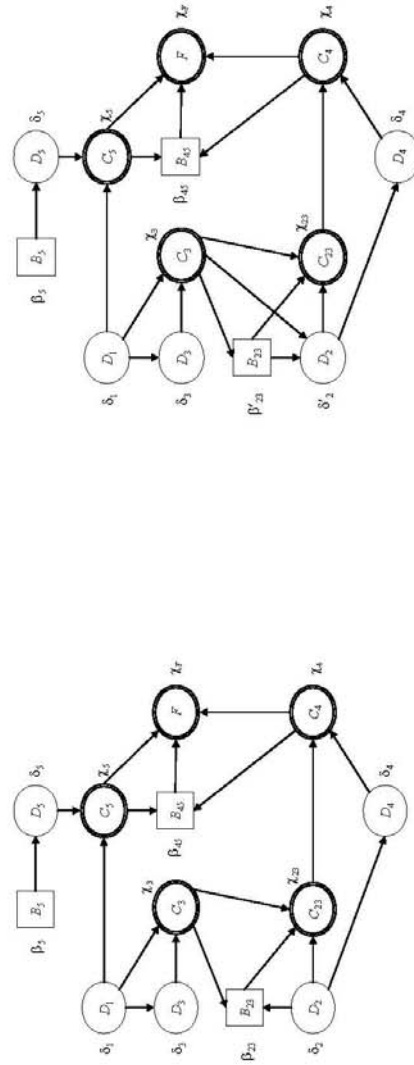
Figure A2: Bayes Net Model

$\delta_2(d_2) = (1, \varphi_{14,3}(d_2))$

$\beta_{23}(b_2, c_3, d_2) = (1, \iota)$ if $c_3 - d_2 \leq 0$

$= (0, \iota)$ if $c_3 - d_2 > 0$

$\beta_{23}(b_3, c_3, d_2) = (0, \iota)$ if $c_3 - d_2 \leq 0$

$= (1, \iota)$ if $c_3 - d_2 > 0$

$$(\delta_2 \otimes \beta_{23})^{-D_2}(b_2, c_3) = \beta'(b_2, c_3) = \left(\int_{-\infty}^{c_3} \varphi_{14,3}(d_2)dd_2, \iota\right)$$

$$(\delta_2 \otimes \beta_{23})^{-D_2}(b_3, c_3) = \beta'(b_3, c_3) = \left(\int_{c_3}^{\infty} \varphi_{14,3}(d_2)dd_2, \iota\right)$$

Let $P(b_2| c_3)$ denote $\int_{-\infty}^{c_3} \varphi_{14,3}(d_2)dd_2$ and $P(b_3| c_3)$

denote $\int_{c_3}^{\infty} \varphi_{14,3}(d_2)dd_2.$

$$((\delta_2 \otimes \beta_{23}) / (\delta_2 \otimes \beta_{23})^{-D_2})(d_2, b_2, c_3) = \delta'_2(d_2, b_2, c_3) =$$

$$(1/P(b_2| c_3), \varphi_{14,3}(d_2))$$

$$((\delta_2 \otimes \beta_{23}) / (\delta_2 \otimes \beta_{23})^{-D_2})(d_2, b_3, c_3) = \delta'_2(d_2, b_3, c_3) =$$

$$(1/P(b_3| c_3), \varphi_{14,3}(d_2))$$

Reversal of arc (C_3, B_{23})

Figure A5: After reversing (C_3, B_{23})

$$\chi'_3 : \quad C_3 = D_1 + D_3$$

$$\beta''_{23}(b_2, d_1, d_3) = \left(\int_{-\infty}^{d_1+d_3} \varphi_{14,3}(d_2)\,dd_2, \ \iota \right)$$

$$\beta''_{23}(b_3, d_1, d_3) = \left(\int_{d_1+d_3}^{\infty} \varphi_{14,3}(d_2)\,dd_2, \ \iota \right)$$

Figure A4: Before reversing (C_3, B_{23})

$$\chi_3 : C_3 = D_1 + D_3$$

$$\beta'_{23}(b_2, c_3) = \left(\int_{-\infty}^{c_3} \varphi_{14,3}(d_2)\,dd_2, \ \iota \right)$$

$$\beta'_{23}(b_3, c_3) = \left(\int_{c_3}^{\infty} \varphi_{14,3}(d_2)\,dd_2, \ \iota \right)$$

Let $P(b_2 | d_1, d_3)$ denote $\displaystyle\int_{-\infty}^{d_1+d_3} \varphi_{14,3}(d_2) dd_2$ and let $P(b_3 | d_1,$

$d_3)$ denote $\displaystyle\int_{d_1+d_3}^{\infty} \varphi_{14,3}(d_2) dd_2.$

Reversal of arc (D_3, B_{23})
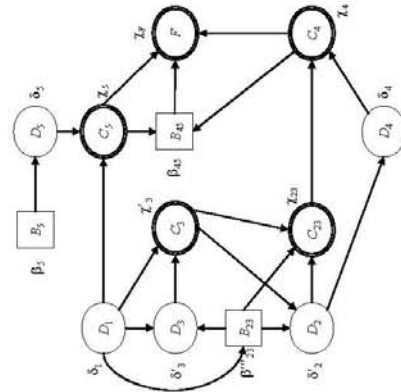
Figure A7: After reversing (D_3, B_{23})

$(\delta_3 \otimes \beta''_{23})^{-D_3}(b_2, d_1) = \beta'''_{23}(b_2, d_1) =$

$(\int\limits_{-\infty}^{\infty} P(b_2| d_1, d_3)\, \varphi_{5+2d_1,\, 2}(d_3)\, dd_3,\ \upsilon)$

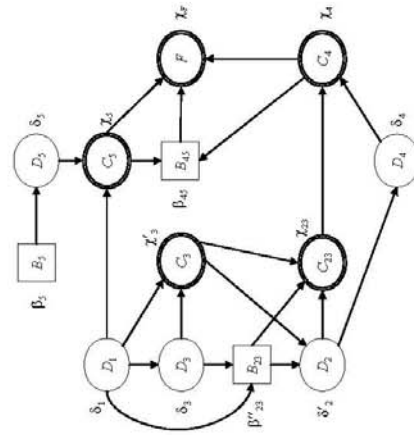$(\delta_3 \otimes \beta''_{23})^{-D_3}(b_3, d_1) = \beta'''_{23}(b_3, d_1) =$

$(\int\limits_{-\infty}^{\infty} P(b_3| d_1, d_3)\, \varphi_{5+2d_1,\, 2}(d_3)\, dd_3,\ \upsilon)$
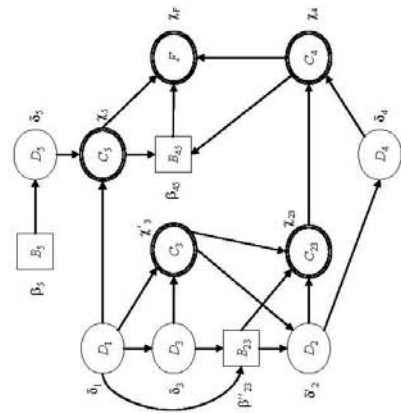
Figure A6: Before reversing (D_3, B_{23})

$\delta_3(d_1, d_3) = (1,\ \varphi_{5+2d_1,\, 2}(d_3))$

$\beta''_{23}(b_2, d_1, d_3) = (P(b_2| d_1, d_3),\ \upsilon)$

$\beta''_{23}(b_3, d_1, d_3) = (P(b_3| d_1, d_3),\ \upsilon)$

$(\delta_3 \otimes \beta''_{23})(b_2, d_1, d_3) = (P(b_2| d_1, d_3),\ \varphi_{5+2d_1,\, 2}(d_3))$

$(\delta_3 \otimes \beta''_{23})(b_3, d_1, d_3) = (P(b_3| d_1, d_3),\ \varphi_{5+2d_1,\, 2}(d_3))$

Let $P(b_2| d_1)$ denote $\int_{-\infty}^{\infty} P(b_2| d_1, d_3)\, \varphi_{5+2d_1, 2}(d_3)dd_3$ and let

$P(b_3| d_1)$ denote $\int_{-\infty}^{\infty} P(b_3| d_1, d_3)\, \varphi_{5+2d_1, 2}(d_3)dd_3$

$$((\delta_3 \otimes \beta''_{23}) / (\delta_3 \otimes \beta''_{23})^{-D_{23}})(d_3, d_1, b_2) = \delta'_3(d_3, d_1, b_2) =$$

$$(P(b_2| d_1, d_3) / P(b_2| d_1),\, \varphi_{5+2d_1, 2}(d_3))$$

$$((\delta_3 \otimes \beta''_{23}) / (\delta_3 \otimes \beta'_{23})^{-D_{23}})(d_3, d_1, b_2) = \delta'_3(d_3, d_1, b_3) =$$

$$(P(b_3| d_1, d_3) / P(b_3| d_1),\, \varphi_{5+2d_1, 2}(d_3))$$

Reversal of arc (D_1, B_{23})
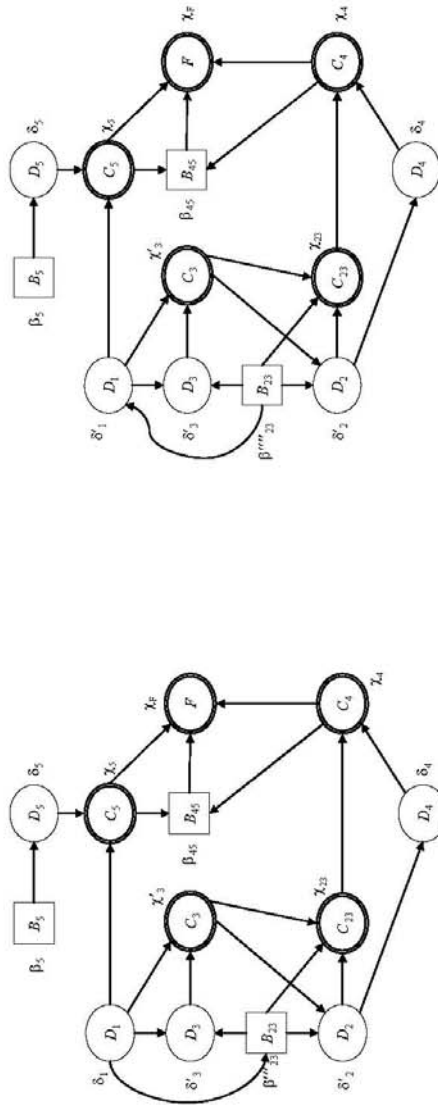
Figure A9: After reversing (D_1, B_{23})

$$(\delta_1 \otimes \beta''''_{23})^{-D_1} = \beta''''_{23}(b_2) = \left(\int_{-\infty}^{\infty} P(b_2 \mid d_1)\varphi_{3,1}(d_1)\, dd_1,\ \iota\right)$$

$$(\delta_1 \otimes \beta''''_{23})^{-D_1} = \beta''''_{23}(b_3) = \left(\int_{-\infty}^{\infty} P(b_3 \mid d_1)\varphi_{3,1}(d_1)\, dd_1,\ \iota\right)$$

Figure A8: Before reversing (D_1, B_{23})

$$\delta_1(d_1) = (1,\ \varphi_{3,1}(d_1))$$

$$\beta'''_{23}(b_2, d_1) = (P(b_2 \mid d_1),\ \iota)$$

$$\beta'''_{23}(b_3, d_1) = (P(b_3 \mid d_1),\ \iota)$$

$$(\delta_1 \otimes \beta'''_{23})\,(b_2, d_1) = (P(b_2 \mid d_1),\ \varphi_{3,1}(d_1))$$

$$(\delta_1 \otimes \beta'''_{23})\,(b_3, d_1) = (P(b_3 \mid d_1),\ \varphi_{3,1}(d_1))$$

Let $P(b_2)$ denote $\int_{-\infty}^{\infty} P(b_2|\,d_1)\varphi_{3,\,1}(d_1)\,dd_1$ and let $P(b_3)$

denote $\int_{-\infty}^{\infty} P(b_3|\,d_1)\varphi_{3,\,1}(d_1)\,dd_1$

$(\delta_1{\otimes}\beta'''_{23})\,/\,(\delta_1{\otimes}\beta'''_{23})^{-D_1} = \delta'_1(d_1,\,b_2) =$

$(P(b_2|\,d_1)\,/\,P(b_2),\,\varphi_{3,\,1}(d_1))$

$(\delta_1{\otimes}\beta'''_{23})\,/\,(\delta_1{\otimes}\beta'''_{23})^{-D_1} = \delta'_1(d_1,\,b_3) =$

$(P(b_3|\,d_1)\,/\,P(b_3),\,\varphi_{3,\,1}(d_1))$

Reversal of arc (C_5, B_{45})
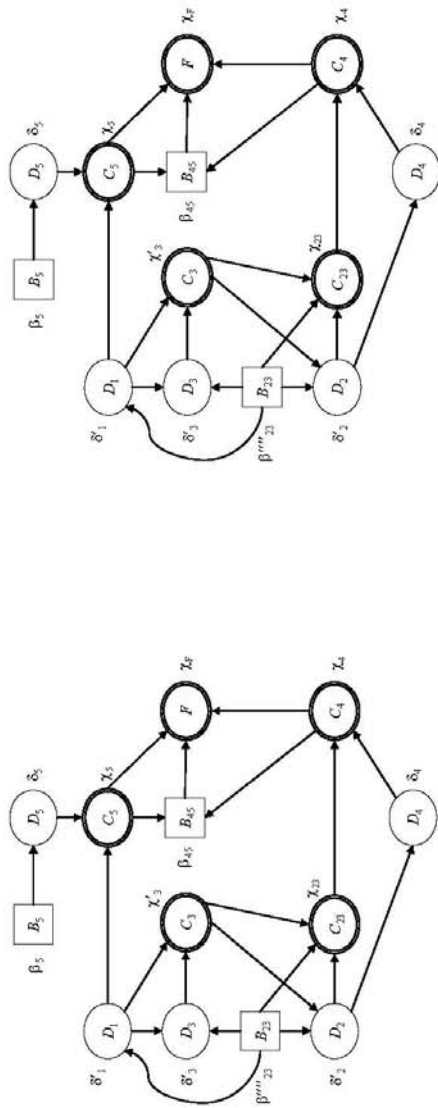
Figure A10: Before reversing (C_5, B_{45})

χ_5: $C_5 = D_1 + D_5$

$$\beta_{45}(b_4, c_4, c_5) = (1, \iota) \qquad \text{if } c_5 - c_4 \leq 0$$
$$= (0, \iota) \qquad \text{if } c_5 - c_4 > 0$$

$$\beta_{45}(b_5, c_4, c_5) = (0, \iota) \qquad \text{if } c_5 - c_4 \leq 0$$
$$= (1, \iota) \qquad \text{if } c_5 - c_4 > 0$$

Figure A11: After reversing (C_5, B_{45})

χ'_5: $C_5 = D_1 + D_5$

$$\beta'_{45}(b_4, d_1, d_5, c_4) = (1, \iota) \qquad \text{if } d_1 + d_5 - c_4 \leq 0$$
$$= (0, \iota) \qquad \text{if } d_1 + d_5 - c_4 > 0$$

$$\beta'_{45}(b_5, d_1, d_5, c_4) = (0, \iota) \qquad \text{if } d_1 + d_5 - c_4 \leq 0$$
$$= (1, \iota) \qquad \text{if } d_1 + d_5 - c_4 > 0$$

Reversal of arc (C_4, B_{45})

Figure A13: After reversing (C_4, B_{45})

$$\beta'_{45}\,(b_5, d_1, d_5, c_4) = (0, \upsilon) \qquad \text{if } d_1 + d_5 - c_4 \leq 0$$
$$= (1, \upsilon) \qquad \text{if } d_1 + d_5 - c_4 > 0$$
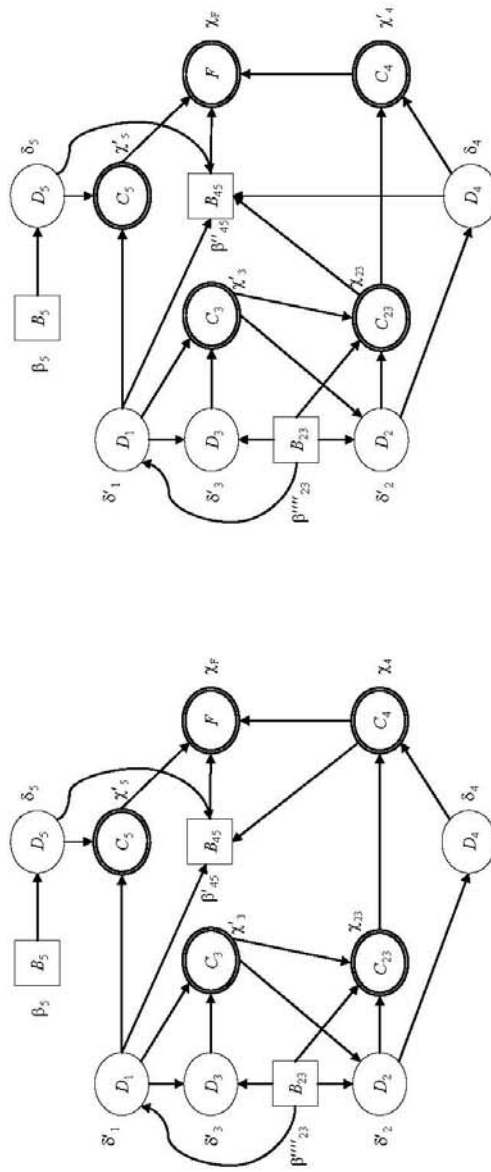
Figure A12: Before reversing (C_4, B_{45})

$$\chi_4: C_4 = C_{23} + D_4$$
$$\beta'_{45}\,(b_4, d_1, d_5, c_4) = (1, \upsilon) \qquad \text{if } d_1 + d_5 - c_4 \leq 0$$
$$= (0, \upsilon) \qquad \text{if } d_1 + d_5 - c_4 > 0$$

$$\chi'_4 : C_4 = C_{23} + D_4$$

$$\beta''_{45}(b_4, d_1, d_5, c_{23}, d_4) = (1, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 \leq 0$$

$$= (0, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 > 0$$

$$\beta''_{45}(b_5, d_1, d_5, c_{23}, d_4) = (0, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 \leq 0$$

$$= (1, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 > 0$$

Reversal of arc (D_4, B_{45})
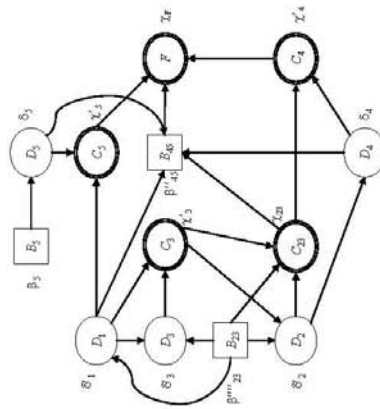
Figure A14: Before reversing (D_4, B_{45})

$$\delta_4(d_2, d_4) = (1, \varphi_{1+d_2, 4}(d_4))$$

$$\beta''_{45}(b_4, d_1, d_5, c_{23}, d_4) = (1, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 \leq 0$$

$$= (0, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 > 0$$

$$\beta''_{45}(b_5, d_1, d_5, c_{23}, d_4) = (0, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 \leq 0$$
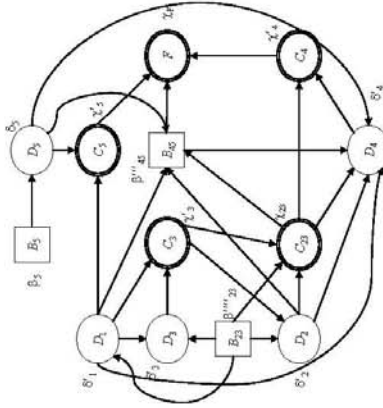
$$= (1, \iota) \text{ if } d_1 + d_5 - c_{23} - d_4 > 0$$

Figure A15: After reversing (D_4, B_{45})

$$(\delta_4 \otimes \beta''_{45})(b_4, d_1, d_5, d_2, c_{23}, d_4) = (1, \varphi_{1+d_2, 4}(d_4))$$

$$\text{if } d_1 + d_5 - c_{23} - d_4 \leq 0$$

$$(\delta_4 \otimes \beta''_{45})(b_5, d_1, d_5, d_2, c_{23}, d_4) =$$

$$(1, \varphi_{1+d_2, 4}(d_4)) \text{ if } d_1 + d_5 - c_{23} - d_4 > 0$$

$$(\delta_4 \otimes \beta''_{45})^{-D_4}(b_4, d_1, d_5, d_2, c_{23}) = \beta'''_{45}(b_4, d_1, d_5, d_2, c_{23}) =$$

$$\left(\int_{-\infty}^{d_1+d_5-c_{23}} \varphi_{1+d_2,4}(d_4)\, dd_4,\ \iota \right)$$

$$(\delta_4 \otimes \beta'_{45})^{-D_4}(b_5, d_1, d_5, d_2, c_{23}) = \beta'''_{45}(b_5, d_1, d_5, d_2, c_{23}) =$$

$$\left(\int_{d_1+d_5-c_{23}}^{\infty} \varphi_{1+d_2,4}(d_4)\, dd_4,\ \iota \right)$$

Let $P(b_4| d_1, d_5, d_2, c_{23})$ denote $\displaystyle\int_{-\infty}^{d_1+d_5-c_{23}} \varphi_{1+d_2,4}(d_4)\, dd_4$ and let

$P(b_5| d_1, d_5, d_2, c_{23})$ denote

$$\int_{d_1+d_5-c_{23}}^{\infty} \varphi_{1+d_2,4}(d_4)\, dd_4 \cdot (\delta_4 \otimes \beta''_{45}) / (\delta_4 \otimes \beta''_{45})^{-D_4} =$$

$$\delta'_4(d_4, b_4, d_1, d_2, c_{23}) = (1 / P(b_4| d_1, d_5, d_2, c_{23}),\ \varphi_{1+d_2,4}(d_4))$$

$$(\delta_4 \otimes \beta''_{45}) / (\delta_4 \otimes \beta''_{45})^{-D_4} = \delta'_4(d_5, b_5, d_1, d_2, c_{23}) =$$

$$(1 / P(b_5| d_1, d_5, d_2, c_{23}),\ \varphi_{1+d_2,4}(d_4))$$

Reversal of arc (C_{23}, B_{45})
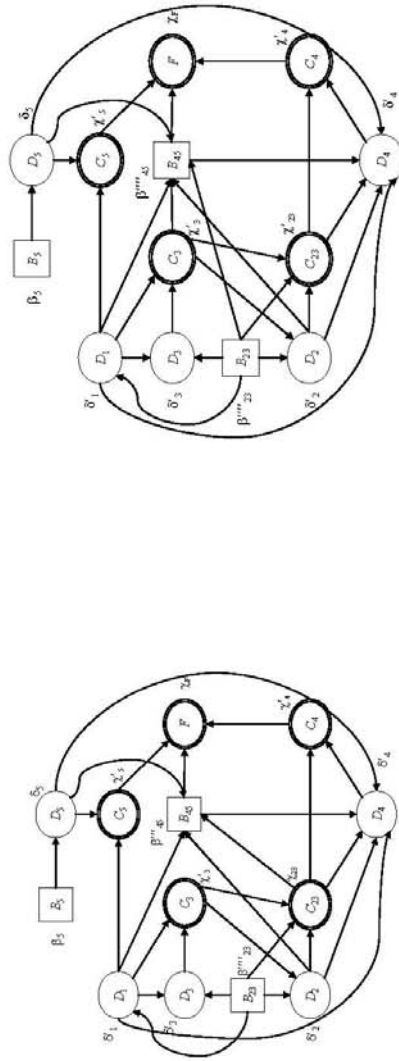
Figure A16: Before reversing (C_{23}, B_{45})

χ_{23}: $\quad C_{23} = D_2 \qquad$ if $B_{23} = b_2$

$\qquad\quad C_{23} = C_3 \qquad$ if $B_{23} = b_3$

$$\beta'''_{45}(b_4, d_1, d_5, d_2, c_{23}) = \left(\int_{-\infty}^{d_1+d_5-c_{23}} \varphi_{1+d_2,4}(d_4)\,dd_4,\ \iota \right)$$

$$\beta'''_{45}(b_5, d_1, d_5, d_2, c_{23}) = \left(\int_{d_1+d_5-c_{23}}^{s} \varphi_{1+d_2,4}(d_4)\,dd_4,\ \iota \right)$$

Figure A17: After reversing (C_{23}, B_{45})

χ'_{23}: $\quad C_{23} = D_2 \qquad$ if $B_{23} = b_2$

$\qquad\quad C_{23} = C_3 \qquad$ if $B_{23} = b_3$

$$\beta''''_{45}(b_4, d_1, d_5, d_2, b_2, c_3) = \left(\int_{-\infty}^{d_1+d_5-d_2} \varphi_{1+d_2,4}(d_4)\,dd_4,\ \iota \right)$$

$$\beta''''_{45}(b_4, d_1, d_5, d_2, b_3, c_3) = \left(\int_{-\infty}^{d_1+d_5-c_3} \varphi_{1+d_2,4}(d_4)\,dd_4,\ \iota \right)$$

$$\beta'''_{45}(b_5, d_1, d_5, d_2, b_2, c_3) = \left(\int_{d_1+d_5-d_2}^{\infty} \varphi_{1+d_2,4}(d_4)\,dd_4, \iota\right)$$

$$\beta'''_{45}(b_2, d_1, d_5, d_2, b_3, c_3) = \left(\int_{d_1+d_5-c_3}^{\infty} \varphi_{1+d_2,4}(d_4)\,dd_4, \iota\right)$$

Let $P(b_4|\ d_1, d_5, d_2, b_2, c_3)$ denote $\int_{-\infty}^{d_1+d_5-d_2} \varphi_{1+d_2,4}(d_4)\,dd_4$ and

let $P(b_4|\ d_1, d_5, d_2, b_3, c_3)$ denote $\int_{-\infty}^{d_1+d_5-c_3} \varphi_{1+d_2,4}(d_4)\,dd_4$. Let

$P(b_5|\ d_1, d_5, d_2, b_2, c_3)$ denote $\int_{d_1+d_5-d_2}^{\infty} \varphi_{1+d_2,4}(d_4)\,dd_4$ and let

$P(b_5|\ d_1, d_5, d_2, b_3, c_3)$ denote $\int_{d_1+d_5-c_3}^{\infty} \varphi_{1+d_2,4}(d_4)\,dd_4$.

Reversal of arc (D_2, B_{45})

Figure A19: After reversing (D_2, B_{45})

$(\delta'_2 \otimes \beta''''_{45})(b_4, d_1, d_5, d_2, b_2, c_3) =$

$(P(b_4|\, d_1, d_5, d_2, \, b_2, c_3) \,/\, P(b_2|\, c_3), \, \varphi_{14,3}(d_2))$

$(\delta'_2 \otimes \beta''''_{45})(b_4, d_1, d_5, d_2, b_3, c_3) =$

$(P(b_4|\, d_1, d_5, d_2, b_3, c_3) \,/\, P(b_3|\, c_3), \, \varphi_{14,3}(d_2))$

$(\delta'_2 \otimes \beta''''_{45})(b_5, d_1, d_5, d_2, b_2, c_3) =$

$(P(b_5|\, d_1, d_5, d_2, b_2, c_3) \,/\, P(b_2|\, c_3), \, \varphi_{14,3}(d_2))$
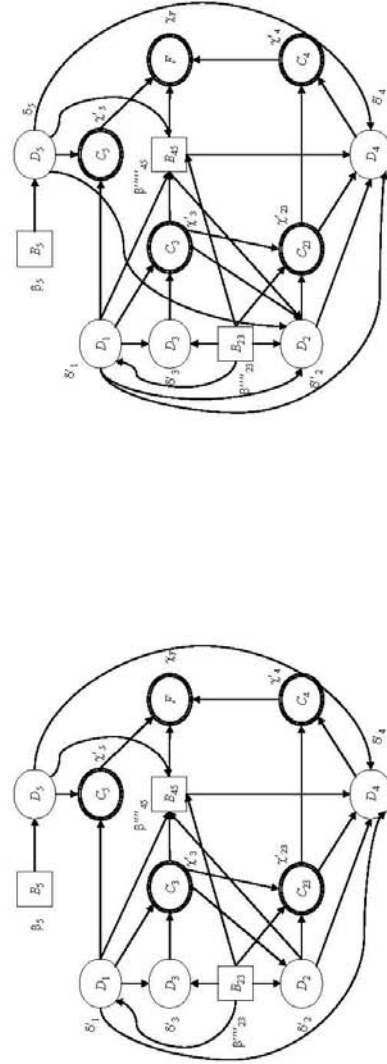
Figure A18: Before reversing (D_2, B_{45})

$\delta'_2(d_2, b_2, c_3) = (1/P(b_2|\, c_3), \, \varphi_{14,3}(d_2))$

$\delta'_2(d_2, b_3, c_3) = (1/P(b_3|\, c_3), \, \varphi_{14,3}(d_2))$

$\beta''''_{45}(b_4, d_1, d_5, d_2, b_2, c_3) = (P(b_4|\, d_1, d_5, d_2, \, b_2, c_3), \, \iota)$

$\beta''''_{45}(b_4, d_1, d_5, d_2, b_3, c_3) = (P(b_4|\, d_1, d_5, d_2, \, b_3, c_3), \, \iota)$

$\beta''''_{45}(b_5, d_1, d_5, d_2, b_2, c_3) = (P(b_5|\, d_1, d_5, d_2, b_2, c_3), \, \iota)$

$\beta''''_{45}(b_5, d_1, d_5, d_2, b_3, c_3) = (P(b_5|\, d_1, d_5, d_2, b_3, c_3), \, \iota)$

$(\delta'_2 \otimes \beta''''_{45})(b_5, d_1, d_5, d_2, b_3, c_3) =$

$(P(b_5 | d_1, d_5, d_2, b_3, c_3) / P(b_3| c_3), \varphi_{14,3}(d_2)$

$(\delta'_2 \otimes \beta''''_{45})^{-D_2}(b_4, d_1, d_5, b_2, c_3) = \beta'''''_{45}(b_4, d_1, d_5, b_2, c_3) =$

$= \int_{-\infty}^{\infty} (P(b_4 | d_1, d_5, d_2, b_2, c_3)\varphi_{14,3}(d_2) / P(b_2| c_3)) \, dd_2, \iota)$

$(\delta'_2 \otimes \beta''''_{45})^{-D_2}(b_4, d_1, d_5, b_3, c_3) = \beta'''''_{45}(b_4, d_1, d_5, b_3, c_3) =$

$\int_{-\infty}^{\infty} (P(b_4 | d_1, d_5, d_2, b_3, c_3)\varphi_{14,3}(d_2) / P(b_3| c_3)) \, dd_2, \iota)$

$(\delta'_2 \otimes \beta'''''_{45})^{-D_2}(b_5, d_1, d_5, b_2, c_3) = \beta'''''_{45}(b_5, d_1, d_5, b_2, c_3) =$

$\int_{-\infty}^{\infty} (P(b_5 | d_1, d_5, d_2, b_2, c_3)\varphi_{14,3}(d_2) / P(b_2| c_3)) \, dd_2, \iota)$

$(\delta'_2 \otimes \beta'''''_{45})^{-D_2}(b_5, d_1, d_5, b_3, c_3) = \beta''''''_{45}(b_5, d_1, d_5, b_3, c_3)$

$= \int_{-\infty}^{\infty} (P(b_5 | d_1, d_5, d_2, b_3, c_3)\varphi_{14,3}/(P(b_3| c_3))(d_2) \, dd_2, \iota)$

Let $P(b_4 | d_1, d_5, d_2, b_2, c_3)$ denote

$\int_{-\infty}^{\infty} P(b_4 | d_1, d_5, d_2, b_2, c_3)\varphi_{14,3}(d_2) / P(b_2| c_3) \, dd_2$, let

$P(b_4 | d_1, d_5, b_2, c_3)$ denote

$\int_{-\infty}^{\infty} P(b_4 | d_1, d_5, d_2, b_3, c_3)\varphi_{14,3}(d_2) / P(b_3| c_3) \, dd_2$, let

$P(b_5 | d_1, d_5, b_2, c_3)$ denote

$\int_{-\infty}^{\infty} (P(b_5 | d_1, d_5, d_2, b_2, c_3)\varphi_{14,3}(d_2) / P(b_2| c_3)) \, dd_2$ and let

$P(b_5 | d_1, d_5, b_3, c_3)$ denote

$\int_{-\infty}^{\infty} (P(b_5 | d_1, d_5, d_2, b_3, c_3)\varphi_{14,3}(d_2) / (P(b_3| c_3)) \, dd_2$

$(\delta'_2 \otimes \beta'''''_{45})^{-D_2} = \delta''_2(d_2, b_4, d_1, d_5, b_2, c_3) =$

$(P(b_4 | d_1, d_5, d_2, b_2, c_3) (P(b_4 | d_1, d_5, b_2, c_3), \varphi_{14,3}(d_2))$

$$(\delta'_2 \otimes \beta''''_{45}) / (\delta'_2 \otimes \beta''''_{45})^{-D_2} = \delta''_2(d_2, b_4, d_1, d_5, b_3, c_3) =$$

$$(P(b_4| d_1, d_5, d_2, b_3, c_3)/ (P(b_3| c_3) P(b_4| d_1, d_5, b_2, c_3), \varphi_{14},$$

$$_3(d_2))$$

$$(\delta'_2 \otimes \beta''''_{45}) / (\delta'_2 \otimes \beta''''_{45})^{-D_2} = \delta''_2(d_2, b_5, d_1, d_5, b_2, c_3) =$$

$$(P(b_5| d_1, d_5, b_2, c_3)/ P(b_2| c_3) P(b_5| d_1, d_5, b_2, c_3), \varphi_{14},$$

$$_3(d_2))$$

$$(\delta'_2 \otimes \beta''''_{45}) / (\delta'_2 \otimes \beta''''_{45})^{-D_2} = \delta''_2(d_2, b_5, d_1, d_5, b_3, c_3) =$$

$$(P(b_5| d_1, d_5, d_2, b_3, c_3)/ P(b_5| d_1, d_5, b_3, c_3) P(b_3| c_3), \varphi_{14},$$

$$_3(d_2))$$

Reversal of arc (C_3, B_{45})

Figure A21: After reversing (C_3, B_{45})

χ''_3: $\quad C_3 = D_1 + D_3$

$\beta''''''_{45}(b_4, d_1, d_5, b_2, d_1, d_3) = (P(b_4 \mid d_1, d_5, b_2, d_3), \iota)$

$\beta''''''_{45}(b_4, d_1, d_5, b_3, d_1, d_3) = (P(b_4 \mid d_1, d_5, b_2, d_3), \iota)$

$\beta''''''_{45}(b_5, d_1, d_5, b_2, d_1, d_3) = (P(b_5 \mid d_1, d_5, b_2, d_3), \iota)$

$\beta''''''_{45}(b_5, d_1, d_5, b_3, d_1, d_3) = (P(b_5 \mid d_1, d_5, b_3, d_3), \iota)$

Figure A20: Before reversing (C_3, B_{45})

χ'_3: $\quad C_3 = D_1 + D_3$

$\beta'''''_{45}(b_4, d_1, d_5, b_2, c_3) = P(b_4 \mid d_1, d_5, b_2, c_3), \iota)$

$\beta'''''_{45}(b_4, d_1, d_5, b_3, c_3) = P(b_4 \mid d_1, d_5, b_2, c_3), \iota)$

$\beta'''''_{45}(b_5, d_1, d_5, b_2, c_3) = P(b_5 \mid d_1, d_5, b_2, c_3), \iota)$

$\beta'''''_{45}(b_5, d_1, d_5, b_3, c_3) = P(b_5 \mid d_1, d_5, b_3, c_3), \iota)$

Reversal of arc (D_3, B_{45})

Figure A23: After reversing (D_3, B_{45})

$\beta''''_{45}(b_5, d_1, d_5, b_3, d_3) = (P(b_3 \mid d_1, d_5, b_3, d_3), \iota)$

$(\delta'_3 \otimes \beta''''_{45})(b_4, d_1, d_5, b_2, d_3) =$

$(P(b_2 \mid d_1, d_3) P(b_4 \mid d_1, d_5, b_2, d_3) / P(b_2 \mid d_1), \varphi_{5+2d_1}, 2(d_3))$

$(\delta'_3 \otimes \beta''''_{45})(b_4, d_1, d_5, b_3, d_3) =$

$(P(b_3 \mid d_1, d_3) P(b_4 \mid d_1, d_5, b_2, d_3) / P(b_3 \mid d_1), \varphi_{5+2d_1}, 2(d_3))$
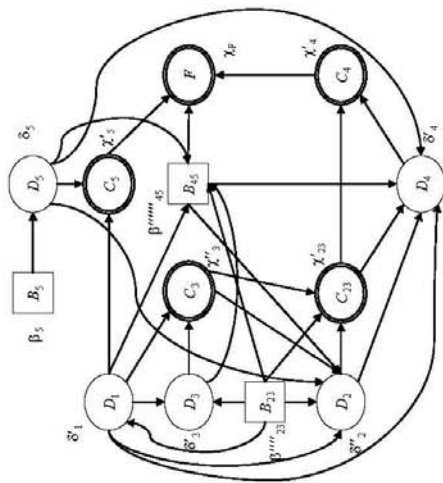
Figure A22: Before reversing (D_3, B_{45})

$\delta'_3(d_3, d_1, b_2) = (P(b_2 \mid d_1, d_3) / P(b_2 \mid d_1), \varphi_{5+2d_1}, 2(d_3))$

$\delta'_3(d_3, d_1, b_3) = (P(b_3 \mid d_1, d_3) / P(b_3 \mid d_1), \varphi_{5+2d_1}, 2(d_3))$

$\beta'''''_{45}(b_4, d_1, d_5, b_2, d_3) = (P(b_4 \mid d_1, d_5, b_2, d_3), \iota)$

$\beta'''''_{45}(b_4, d_1, d_5, b_3, d_3) = (P(b_4 \mid d_1, d_5, b_2, d_3), \iota)$

$\beta'''''_{45}(b_5, d_1, d_5, b_2, d_3) = (P(b_5 \mid d_1, d_5, b_2, d_3), \iota)$

$$(\delta'_3 \otimes \beta''''''_{45})(b_5, d_1, d_5, b_2, d_3) =$$

$$(P(b_2| d_1, d_3) P(b_5| d_1, d_5, b_2, d_3) / P(b_2| d_1), \varphi_{5+2d_1, 2}(d_3))$$

$$(\delta'_3 \otimes \beta''''''_{45})(b_5, d_1, d_5, b_3, d_3) =$$

$$(P(b_3| d_1, d_3) P(b_5| d_1, d_5, b_3, d_3) / P(b_3| d_1), \varphi_{5+2d_1, 2}(d_3))$$

$$(\delta'_3 \otimes \beta''''''_{45})^{-D_3}(b_4, d_1, d_5, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_2| d_1, d_3) P(b_4| d_1, d_5, b_2, d_3) / P(b_2| d_1) \varphi_{5+2d_1, 2}(d_3))$$

$$dd_3, \iota)$$

Let $P(b_4| d_1, d_5, b_2)$ denote

$$(\delta'_3 \otimes \beta''''''_{45})^{-D_3}(b_4, d_1, d_5, b_3) = \beta'''''''_{45}(b_4, d_1, d_5, b_3) =$$

$$\int_{-\infty}^{\infty} (P(b_3| d_1, d_3) P(b_4| d_1, d_5, b_2, d_3) / (P(b_3| d_1) \varphi_{5+2d_1, 2}(d_3))$$

$$dd_3, \iota)$$

$$(\delta'_3 \otimes \beta''''''_{45})^{-D_3}(b_5, d_1, d_5, b_2) = \beta'''''''_{45}(b_5, d_1, d_5, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_2| d_1, d_3) P(b_5| d_1, d_5, b_2, d_3)) / P(b_2| d_1)\varphi_{5+2d_1, 2}(d_3))$$

$$dd_3, \iota)$$

$$(\delta'_3 \otimes \beta''''''_{45})^{-D_3}(b_5, d_1, d_5, b_3) = \beta'''''''_{45}(b_5, d_1, d_5, b_3) =$$

$$\int_{-\infty}^{\infty} (P(b_3| d_1, d_3) P(b_5| d_1, d_5, b_3, d_3) / P(b_3| d_1)\varphi_{5+2d_1, 2}(d_3))$$

$$dd_3, \iota)$$

Let $P(b_4| d_1, d_5, b_2)$ denote

$$\int_{-\infty}^{\infty} (P(b_2| d_1, d_3) P(b_4| d_1, d_5, b_2, d_3) / P(b_2| d_1) \varphi_{5+2d_1, 2}(d_3))$$

dd_3, let $P(b_4| d_1, d_5, b_3)$ denote $\int_{-\infty}^{\infty} (P(b_3| d_1, d_3) P(b_4| d_1, d_5,$

$b_2, d_3) / P(b_3| d_1) \varphi_{5+2d_1, 2}(d_3)) dd_3$, let $P(b_5| d_1, d_5, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_2| d_1, d_3) P(b_5| d_1, d_5, b_2, d_3)) / P(b_2| d_1)\varphi_{5+2d_1, 2}(d_3))$

dd_3 and let $P(b_5| d_1, d_5, b_3)$ denote $\int_{-\infty}^{\infty} (P(b_3| d_1, d_3) P(b_5| d_1,$

$d_5, b_3, d_3) / P(b_3| \mathbf{d_1})\varphi_{5+2d_1, 2}(d_3))\, dd_3.$

$(\delta'_3 \otimes \beta'''''_{45}) / (\delta'_3 \otimes \beta''''''_{45})^{-D_3} = \delta''_3(d_3, b_4, d_1, d_5, b_2) =$

$(P(b_2| d_1, d_3) P(b_4| d_1, d_5, b_2, d_3) / (P(b_2| d_1) P(b_4| d_1, d_5, b_2)).$

$\varphi_{5+2d_1, 2}(d_3))$

$(\delta'_3 \otimes \beta''''''_{45}) / (\delta'_3 \otimes \beta''''''_{45})^{-D_3} = \delta''_3(d_3, b_4, d_1, d_5, b_3) =$

$(P(b_3| d_1, d_3) P(b_4| d_1, d_5, b_2, d_3) / (P(b_3| d_1) P(b_4| d_1, d_5, b_3)).$

$\varphi_{5+2d_1, 2}(d_3))$

$(\delta'_3 \otimes \beta''''''_{45}) / (\delta'_3 \otimes \beta''''''_{45})^{-D_3} = \delta''_3(d_3, b_5, d_1, d_5, b_2) =$

$(P(b_2| d_1, d_3) P(b_5| d_1, d_5, b_2, d_3) / (P(b_2| d_1) P(b_5| d_1, d_5, b_2))$

$, \varphi_{5+2d_1, 2}(d_3))$

$(\delta'_3 \otimes \beta''''''_{45}) / (\delta'_3 \otimes \beta''''''_{45})^{-D_3} = \delta''_3(d_3, b_5, d_1, d_5, b_3)$

$=(P(b_3| d_1, d_3) P(b_5| d_1, d_5, b_3, d_3) / (P(b_3| d_1) P(b_5| d_1, d_5,$

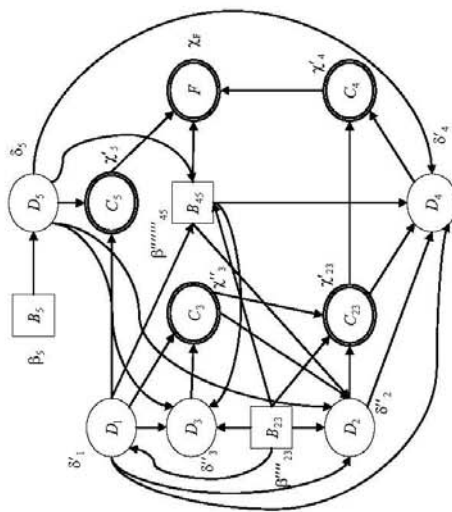$b_3)) , \varphi_{5+2d_1, 2}(d_3)$

Reversal of arc (D_1, B_{45})

Figure A25: After reversing (D_1, B_{45})

$\beta''''''_{45}(b_5, d_1, d_5, b_3) = (P(b_5 | d_1, d_5, b_3), \iota)$

$(\delta'_1 \otimes \beta''''''_{45})(b_4, d_1, d_5, b_2) =$

$(P(b_2 | d_1)P(b_4 | d_1, d_5, b_2) / P(b_2), \varphi_{3,1}(d_1))$

$(\delta'_1 \otimes \beta''''''_{45})(b_4, d_1, d_5, b_3) = (P(b_3 | d_1)P(b_4 | d_1, d_5, b_3) / P(b_3), \varphi_{3,1}(d_1))$

Figure A24: Before reversing (D_1, B_{45})

$\delta'_1(d_1, b_2) = (P(b_2 | d_1) / P(b_2), \varphi_{3,1}(d_1))$

$\delta'_1(d_1, b_3) = (P(b_3 | d_1) / P(b_3), \varphi_{3,1}(d_1))$

$\beta''''''_{45}(b_4, d_1, d_5, b_2) = (P(b_4 | d_1, d_5, b_2), \iota)$

$\beta''''''_{45}(b_4, d_1, d_5, b_3) = (P(b_4 | d_1, d_5, b_3), \iota)$

$\beta''''''_{45}(b_5, d_1, d_5, b_2) = (P(b_5 | d_1, d_5, b_2), \iota)$

$$(\delta'_1 \otimes \beta''''''_{45})(b_5, d_1, d_5, b_2) =$$

$$(P(b_2 | d_1)P(b_5 | d_1, d_5, b_2) / P(b_2)\,\varphi_{3,1}(d_1))$$

$$(\delta'_1 \otimes \beta''''''_{45})(b_5, d_1, d_5, b_3) =$$

$$(P(b_3 | d_1)P(b_5 | d_1, d_5, b_3) / P(b_3)\,\varphi_{3,1}(d_1))$$

$$(\delta'_1 \otimes \beta'''''''_{45})^{-D_1}(b_4, d_5, b_2) =$$

$$\left(\int_{-\infty}^{\infty} (P(b_2 | d_1)P(b_4 | d_1, d_5, b_2) / P(b_2)\varphi_{3,1}(d_1))dd_1, \iota\right)$$

$$(\delta'_1 \otimes \beta'''''''_{45})^{-D_1}(b_4, d_5, b_3) = \beta'''''''_{45}(b_4, d_5, b_3) =$$

$$\left(\int_{-\infty}^{\infty} (P(b_3 | d_1)P(b_4 | d_1, d_5, b_3) / P(b_3)\varphi_{3,1}(d_1))dd_1, \iota\right)$$

$$(\delta'_1 \otimes \beta'''''''_{45})^{-D_1}(b_5, d_5, b_2) =$$

$$\left(\int_{-\infty}^{\infty} (P(b_2 | d_1)P(b_5 | d_1, d_5, b_2) / P(b_2)\varphi_{3,1}(d_1))dd_1, \iota\right)$$

$$(\delta'_1 \otimes \beta'''''''_{45})^{-D_1}(b_5, d_5, b_3) = \beta'''''''_{45}(b_5, d_5, b_3) =$$

$$\left(\int_{-\infty}^{\infty} (P(b_3 | d_1) \text{ Let } P(b_4 | d_5, b_2) \text{ denote}\right.$$

$$\int_{-\infty}^{\infty} (P(b_2 | d_1)P(b_4 | d_1, d_5, b_2) / P(b_2)\varphi_{3,1}(d_1))dd_1, \text{ let}$$

$$P(b_4 | d_5, b_3) \text{ denote}$$

$$\int_{-\infty}^{\infty} (P(b_3 | d_1)P(b_4 | d_1, d_5, b_3) / P(b_3)\varphi_{3,1}(d_1))dd_1, \text{ let}$$

$$P(b_5 | d_5, b_2) \text{ denote}$$

$$\int_{-\infty}^{\infty} (P(b_2 | d_1)P(b_5 | d_1, d_5, b_2) / P(b_2)\varphi_{3,1}(d_1))dd_1 \text{ and let}$$

$$P(b_5 | d_5, b_3) \text{ denote}$$

$$\int_{-\infty}^{\infty} (P(b_3 | d_1)P(b_5 | d_1, d_5, b_3) / P(b_3)\varphi_{3,1}(d_1))dd_1.$$

$$(\delta'_1 \otimes \beta'''''''_{45}) / (\delta'_1 \otimes \beta'''''''_{45})^{-D_1} = \delta''_1(d_1, b_4, d_5, b_2) =$$

$$(P(b_2 | d_1)P(b_4 | d_1, d_5, b_2) / (P(b_4 | d_5, b_2) P(b_2)\,\varphi_{3,1}(d_1))$$

$$(\delta'_1 \otimes \beta'''''''_{45}) \, / \, (\delta'_1 \otimes \beta'''''''_{45})^{-D_1} = \delta''_1(d_1, b_4, d_5, b_3) =$$

$$(P(b_3|\, d_1)P(b_4|\, d_1, d_5, b_3)\, / \, (P(b_3)(P(b_4|\, d_5, b_3)), \varphi_{3,1}(d_1))$$

$$(\delta'_1 \otimes \beta'''''''_{45}) \, / \, (\delta'_1 \otimes \beta'''''''_{45})^{-D_1} = \delta''_1(d_1, b_5, d_5, b_2) =$$

$$((P(b_2|\, d_1)P(b_5|\, d_1, d_5, b_2)\, / \, (P(b_5|\, d_5, b_2)P(b_2)), \varphi_{3,1}(d_1))$$

$$(\delta'_1 \otimes \beta'''''''_{45}) \, / \, (\delta'_1 \otimes \beta'''''''_{45})^{-D_1} = \delta'_1(d_1, b_5, d_5, b_3) =$$

$$((P(b_3|\, d_1)P(b_5|\, d_1, d_5, b_3)\, / \, P(b_5|\, d_5, b_3)P(b_3)\varphi_{3,1}(d_1))$$

$$P(b_5|\, d_1, d_5, b_3)\, / \, P(b_3)\varphi_{3,1}(d_1))\, dd_1, \iota)$$

Reversal of arc (D_5, B_{45})

Figure A27: After reversing (D_5, B_{45})
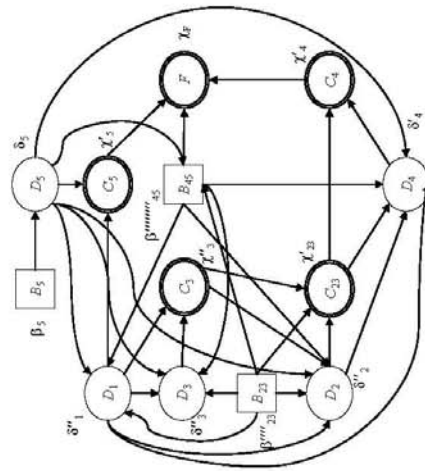
Figure A26: Before reversing (D_5, B_{45})

$\beta''''''_{45}(b_4, d_5, b_2) = (P(b_4| d_5, b_2), \iota)$

$\beta''''''_{45}(b_4, d_5, b_3) = (P(b_4| d_5, b_3), \iota)$

$\beta''''''_{45}(b_5, d_5, b_2) = (P(b_5| d_5, b_2), \iota)$

$\beta''''''_{45}(b_5, d_5, b_3) = (P(b_5| d_5, b_3), \iota)$

$\delta_5(b_{51}, d_5) = (1, \Phi_{0.032, \ 0.0112}(d_5))$

$\delta_5(b_{52}, d_5) = (1, \Phi_{0.143, \ 0.0482}(d_5))$

$\delta_5(b_{53}, d_5) = (1, \Phi_{0.415, \ 0.1382}(d_5))$

$\delta_5(b_{54}, d_5) = (1, \Phi_{1.014, \ 0.3382}(d_5))$

$\delta_5(b_{55}, d_5) = (1, \Phi_{2.300, \ 0.7672}(d_5))$

$$\beta'''''_{45}(b_{51}, b_4, d_5, b_2) =$$

$$(P(b_4| d_5, b_2), \varphi_{0.032,\, 0.0112}(d_5))$$

$$\beta''''''_{45}(b_{52}, b_4, d_5, b_2) =$$

$$(P(b_4| d_5, b_2), \varphi_{0.143,\, 0.0482}(d_5)$$

$$\beta'''''''_{45}(b_{53}, b_4, d_5, b_2) =$$

$$(P(b_4| d_5, b_2), \varphi_{0.415,\, 0.1382}(d_5)$$

$$\beta'''''''_{45}(b_{54}, b_4, d_5, b_2) =$$

$$(P(b_4| d_5, b_2), \varphi_{1.014,\, 0.3382}(d_5)$$

$$\beta'''''''_{45}(b_{55}, b_4, d_5, b_2) =$$

$$(P(b_4| d_5, b_2), \varphi_{2.300,\, 0.7672}(d_5)$$

$$(\delta_5 \otimes \beta'''''')^{-D_5}_{45}(b_4, b_{51}, b_2) = \beta'''''''_{45}(b_4, b_{51}, b_2) =$$

$$(\int_{-\infty}^{\infty} (P(b_4| d_5, b_2)\varphi_{0.032,\, 0.0112}(d_5)) dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''')^{-D_5}_{45}(b_4, b_{52}, b_2) = \beta''''''_{45}(b_4, b_{52}, b_2) =$$

$$(\int_{-\infty}^{\infty} P(b_4| d_5, b_2)\, \varphi_{0.143,\, 0.0482}(d_5)) dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''')^{-D_5}_{45}(b_4, b_{53}, b_2) = \beta'''''''_{45}(b_4, b_{53}, b_2) =$$

$$(\int_{-\infty}^{\infty} P(b_4| d_5, b_2)\, \varphi_{0.415,\, 0.1382}(d_5)) dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''')^{-D_5}_{45}(b_4, b_{54}, b_2) = \beta'''''''_{45}(b_4, b_{54}, b_2) =$$

$$(\int_{-\infty}^{\infty} P(b_4| d_5, b_2)\, \varphi_{1.014,\, 0.3382}(d_5)) dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''')^{-D_5}_{45}(b_4, b_{55}, b_2) = \beta'''''''_{45}(b_4, b_{55}, b_2) =$$

$$\int_{-\infty}^{\infty} P(b_4| d_5, b_2)\, \varphi_{2.300,\, 0.7672}(d_5)) dd_5, \iota)$$

Let $P(b_4| b_{51}, b_2)$ denote

$$\int_{-\infty}^{\infty} (P(b_4| d_5, b_2)\varphi_{0.032,\, 0.0112}(d_5)) dd_5.$$

Let $P(b_4|\ b_{52},\ b_2)$ denote $\int_{-\infty}^{\infty} P(b_4|\ d_5,\ b_2)\ \varphi_{0.143,\ 0.0482}\ (d_5))dd_5.$

Let $P(b_4|\ b_{53},\ b_2)$ denote $\int_{-\infty}^{\infty} P(b_4|\ d_5,\ b_2)\ \varphi_{0.415,\ 0.1382}\ (d_5))dd_5$

Let $P(b_4|\ b_{54},\ b_2)$ denote

$\int_{-\infty}^{\infty} P(b_4|\ d_5,\ b_2)\ \varphi_{1.014,\ 0.3382}\ (d_5))\ dd_5.$

Let $P(b_4|\ b_{55},\ b_2)$ denote $\int_{-\infty}^{\infty} P(b_4|\ d_5,\ b_2)\ \varphi_{2.300,\ 0.7672}\ (d_5))\ dd_5.$

$(\delta_5 \otimes \beta''''''''_{45})\ /\ (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5\ (d_5,\ b_4,\ b_{51},\ b_2) =$

$(P(b_4|\ d_5,\ b_2)\ /\ P(b_4|\ b_{51},\ b_2)),\ \varphi_{0.032,\ 0.0112}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})\ /\ (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta_5\ (d_5,\ b_4,\ b_{52},\ b_2) =$

$(P(b_4|\ d_5,\ b_2)\ /\ P(b_4|\ b_{52},\ b_2),\ \varphi_{0.143,\ 0.0482}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})\ /\ (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5\ (d_5,\ b_4,\ b_{53},\ b_2) =$

$((P(b_4|\ d_5,\ b_2)\ /\ P(b_4|\ b_{53},\ b_2),\ \varphi_{0.415,\ 0.1382}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})\ /\ (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5,\ b_4,\ b_{54},\ b_2) =$

$(P(b_4|\ d_5,\ b_2)\ /\ P(b_4|\ b_{54},\ b_2),\ \varphi_{1.014,\ 0.3382}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})\ /\ (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5,\ b_4,\ b_{55},\ b_2) =$

$(P(b_4|\ d_5,\ b_2)\ /\ P(b_4|\ b_{55},\ b_2),\ \varphi_{2.300,\ 0.7672}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})(b_{51},\ b_4,\ d_5,\ b_3) =$

$(P(b_4|\ d_5,\ b_3),\ \varphi_{0.143,\ 0.0482}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})(b_{53},\ b_4,\ d_5,\ b_3) =$

$(P(b_4|\ d_5,\ b_3),\ \varphi_{0.415,\ 0.1382}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})(b_{54},\ b_4,\ d_5,\ b_3) =$

$(P(b_4|\ d_5,\ b_3),\ \varphi_{1.014,\ 0.3382}\ (d_5))$

$(\delta_5 \otimes \beta''''''''_{45})(b_{55},\ b_4,\ d_5,\ b_3) =$

$(P(b_4|\ d_5,\ b_3),\ \varphi_{2.300,\ 0.7672}\ (d_5))$

Let $P(b_4|b_{51}, b_3)$ denote $\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\,\varphi_{0.032,\,0.0112}(d_5)\,dd_5$

$(\delta_5\otimes\beta''''''''_{45})^{-D_5}(b_{51}, b_4, b_3) = \beta''''''''_{45}(b_4, b_{51}, b_3) =$

$\left(\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\,\varphi_{0.032,\,0.0112}(d_5)\,dd_5,\ \iota\right)$

Let $P(b_4|b_{52}, b_3)$ denote $\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{0.143,\,0.0482}(d_5)\,dd_5$

$(\delta_5\otimes\beta''''''''_{45})^{-D_5}(b_{52}, b_4, b_3) = \beta''''''''_{45}(b_4, b_{52}, b_3) =$

$\left(\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{0.143,\,0.0482}(d_5)\,dd_5,\ \iota\right)$

Let $P(b_4|b_{53}, b_3)$ denote $\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{0.415,\,0.1382}(d_5)\,dd_5$

$(\delta_5\otimes\beta''''''''_{45})^{-D_5}(b_{53}, b_4, b_3) = \beta''''''''_{45}(b_4, b_{53}, b_3) =$

$\left(\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{0.415,\,0.1382}(d_5)\,dd_5,\ \iota\right)$

Let $P(b_4|b_{54}, b_3)$ denote $\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{1.014,\,0.3382}(d_5)\,dd_5$

$(\delta_5\otimes\beta''''''''_{45})^{-D_5}(b_{54}, b_4, b_3) = \beta''''''''_{45}(b_4, b_{54}, b_3) =$

$\left(\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{1.014,\,0.3382}(d_5)\,dd_5,\ \iota\right)$

Let $P(b_4|b_{55}, b_3)$ denote $\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{2.300,\,0.7672}(d_5)\,dd_5$.

$(\delta_5\otimes\beta''''''''_{45})^{-D_5}(b_{55}, b_4, b_3) = \beta''''''''_{45}(b_4, b_{55}, b_3) =$

$\left(\int_{-\infty}^{\infty} P(b_4|d_5, b_3)\varphi_{2.300,\,0.7672}(d_5)\,dd_5,\ \iota\right)$

$(\delta_5\otimes\beta''''''''_{45})/(\delta_5\otimes\beta''''''''_{45})^{-D_5} = \delta'_5(d_5, b_4, b_{51}, b_3) =$

$(P(b_4|d_5, b_3)/P(b_4|b_{51}, b_3)),\ \varphi_{0.032,\,0.0112}(d_5))$

$(\delta_5\otimes\beta''''''''_{45})/(\delta_5\otimes\beta''''''''_{45})^{-D_5} = \delta'_5(d_5, b_4, b_{52}, b_3) =$

$(P(b_4|d_5, b_3)/P(b_4|b_{52}, b_3),\ \varphi_{0.143,\,0.0482}(d_5))$

$(\delta_5\otimes\beta''''''''_{45})/(\delta_5\otimes\beta''''''''_{45})^{-D_5} = \delta'_5(d_5, b_4, b_{53}, b_3) =$

$(P(b_4|d_5, b_3)/P(b_4|b_{53}, b_3),\ \varphi_{0.415,\,0.1382}(d_5))$

$$(\delta_5 \otimes \beta''''''_{45})^{-D_5}(b_{51}, b_5, b_2) = \beta'''''''_{45}(b_5, b_{51}, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2)\, \varphi_{0.032,\, 0.0112}(d_5)) dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{52}, b_5, b_2) = \beta'''''''_{45}(b_5, b_{52}, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2)\, \varphi_{0.143,\, 0.0482}(d_5)\, dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{53}, b_5, b_2) = \beta'''''''_{45}(b_5, b_{53}, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2)\, \varphi_{0.415,\, 0.1382}(d_5)\, dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{54}, b_5, b_2) = \beta'''''''_{45}(b_5, b_{54}, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2)\, \varphi_{1.014,\, 0.3382}(d_5)\, dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{55}, b_5, b_2) = \beta'''''''_{45}(b_5, b_{55}, b_2) =$$

$$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2)\, \varphi_{2.300,\, 0.7672}(d_5)\, dd_5, \iota)$$

$$(\delta_5 \otimes \beta'''''''_{45}) / (\delta_5 \otimes \beta'''''''_{45})^{-D_5} = \delta'_5(d_5, b_4, b_{54}, b_3) =$$

$$(P(b_4| d_5, b_3) / P(b_4| b_{54}, b_3), \varphi_{1.014,\, 0.3382}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45}) / (\delta_5 \otimes \beta'''''''_{45})^{-D_5} = \delta'_5(d_5, b_4, b_{55}, b_3) =$$

$$(P(b_4| d_5, b_3) / P(b_4| b_{55}, b_3), \varphi_{2.300,\, 0.7672}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{51}, b_5, d_5, b_2) =$$

$$(P(b_5| d_5, b_2), \varphi_{0.032,\, 0.0112}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{52}, b_5, d_5, b_2) =$$

$$(P(b_5| d_5, b_2), \varphi_{0.143,\, 0.0482}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{53}, b_5, d_5, b_2) =$$

$$(P(b_5| d_5, b_2), \varphi_{0.415,\, 0.1382}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{54}, b_5, d_5, b_2) =$$

$$(P(b_5| d_5, b_2), \varphi_{1.014,\, 0.3382}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{55}, b_5, d_5, b_2) =$$

$$(P(b_5| d_5, b_2), \varphi_{2.300,\, 0.7672}(d_5))$$

$(\delta_5 \otimes \beta''''''''_{45}) / (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5, b_5, b_{51}, b_2) =$

$(P(b_5| d_5, b_2) / P((b_5| b_{51}, b_2), \varphi_{0.032, 0.0112} (d_5))$

$(\delta_5 \otimes \beta''''''''_{45}) / (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5, b_5, b_{52}, b_2) =$

$(P(b_5| d_5, b_2) / P((b_5| b_{52}, b_2), \varphi_{0.143, 0.0482} (d_5))$

$(\delta_5 \otimes \beta''''''''_{45}) / (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5, b_5, b_{53}, b_2) =$

$(P(b_5| d_5, b_2) / P((b_5| b_{53}, b_2), \varphi_{0.415, 0.1382} (d_5))$

$(\delta_5 \otimes \beta''''''''_{45}) / (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5, b_5, b_{54}, b_2) =$

$(P(b_5| d_5, b_2) /$

$P((b_5| b_{54}, b_2), \varphi_{1.014, 0.3382} (d_5))$

$(\delta_5 \otimes \beta''''''''_{45}) / (\delta_5 \otimes \beta''''''''_{45})^{-D_5} = \delta'_5 (d_5, b_5, b_{55}, b_2) = (P(b_5|$

$d_5, b_2) / P((b_5| b_{55}, b_2), \varphi_{2.300, 0.7672} (d_5))$

$(\delta_5 \otimes \beta''''''''_{45}) (b_{51}, b_5, d_5, b_3) =$

$(P(b_5| d_5, b_3), \varphi_{0.032, 0.0112} (d_5))$

Let $P((b_5| b_{51}, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2) \varphi_{0.032, 0.0112} (d_5) dd_5.$

Let $P((b_5| b_{52}, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2) \varphi_{0.143, 0.0482} (d_5) dd_5.$

Let $P((b_5| b_{53}, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2) \varphi_{0.415, 0.1382} (d_5) dd_5.$

Let $P((b_5| b_{54}, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2) \varphi_{1.014, 0.3382} (d_5) dd_5.$

Let $P((b_5| b_{55}, b_2)$ denote

$\int_{-\infty}^{\infty} (P(b_5| d_5, b_2) \varphi_{2.300, 0.7672} (d_5) dd_5$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{52}, b_5, d_5, b_3) =$$

$$(P(b_5 \mid d_5, b_3), \varphi_{0.143,\,0.0482}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{53}, b_5, d_5, b_3) =$$

$$(P(b_5 \mid d_5, b_3), \varphi_{0.415,\,0.1382}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{54}, b_5, d_5, b_3) =$$

$$(P(b_5 \mid d_5, b_3), \varphi_{1.014,\,0.3382}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})(b_{55}, b_5, d_5, b_3) =$$

$$(P(b_5 \mid d_5, b_3), \varphi_{2.300,\,0.7672}(d_5))$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{51}, b_5, b_3) = \beta''''''''_{45}(b_5, b_{51}, b_3) =$$

$$\left(\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.032,\,0.0112}(d_5)\, dd_5,\ \iota \right)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{52}, b_5, b_3) = \beta''''''''_{45}(b_5, b_{52}, b_3) =$$

$$\left(\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.143,\,0.0482}(d_5)\, dd_5,\ \iota \right)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{53}, b_5, b_3) = \beta''''''''_{45}(b_5, b_{53}, b_3) =$$

$$\left(\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.415,\,0.1382}(d_5)\, dd_5,\ \iota \right)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{54}, b_5, b_3) = \beta''''''''_{45}(b_5, b_{54}, b_3) =$$

$$= \int_{-\infty}^{\infty} \left(P(b_5 \mid d_5, b_3)\, \varphi_{1.014,\,0.3382}(d_5)\, dd_5,\ \iota \right)$$

$$(\delta_5 \otimes \beta'''''''_{45})^{-D_5}(b_{55}, b_5, b_3) = \beta''''''''_{45}(b_5, b_{55}, b_3) =$$

$$= \int_{-\infty}^{\infty} \left(P(b_5 \mid d_5, b_3)\, \varphi_{2.300,\,0.7672}(d_5)\, dd_5,\ \iota \right)$$

Let $P(b_{51} \mid b_5, b_3)$ denote $\displaystyle\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.032,\,0.0112}(d_5)\, dd_5$.

Let $P(b_{52} \mid b_5, b_3)$ denote $\displaystyle\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.143,\,0.0482}(d_5)\, dd_5$.

Let $P(b_{53} \mid b_5, b_3)$ denote $\displaystyle\int_{-\infty}^{\infty} P(b_5 \mid d_5, b_3)\, \varphi_{0.415,\,0.1382}(d_5)\, dd_5$.

Let $P(b_{54}| b_5, b_3)$ denote $\int_{-\infty}^{\infty} (P(b_5| d_5, b_3) \, \varphi_{1.014,\ 0.3382}\ (d_5)\ dd_5.$

Let $P(b_{55}| b_5, b_3)$ denote $\int_{-\infty}^{\infty} (P(b_5| d_5, b_3) \, \varphi_{2.300,\ 0.7672}\ (d_5)\ dd_5.$

$(\delta_5 {\otimes} \beta''''''''_{45}) / (\delta_5 {\otimes} \beta''''''''_{45})^{-D_5} = \delta'_5(d_5,\ b_5,\ b_{51},\ b_3) =$

$(P(b_5| d_5, b_3) / P(b_{51}| b_5, b_3).\ \varphi_{0.032,\ 0.0112}\ (d_5))$

$(\delta_5 {\otimes} \beta''''''''_{45}) / (\delta_5 {\otimes} \beta''''''''_{45})^{-D_5} = \delta'_5(d_5,\ b_5,\ b_{52},\ b_3) =$

$(P(b_5| d_5, b_3) / P(b_{52}| b_5, b_3).\ \varphi_{0.143,\ 0.0482}\ (d_5))$

$(\delta_5 {\otimes} \beta''''''''_{45}) / (\delta_5 {\otimes} \beta''''''''_{45})^{-D_5} = \delta'_5(d_5,\ b_5,\ b_{53},\ b_3) =$

$(P(b_5| d_5, b_3) / P(b_{53}| b_5, b_3).\ \varphi_{0.415,\ 0.1382}\ (d_5))$

$(\delta_5 {\otimes} \beta''''''''_{45}) / (\delta_5 {\otimes} \beta''''''''_{45})^{-D_5} = \delta'_5(d_5,\ b_5,\ b_{54},\ b_3) =$

$(P(b_5| d_5, b_3) / P(b_{54}| b_5, b_3).\ \varphi_{1.014,\ 0.3382}\ (d_5))$

$(\delta_5 {\otimes} \beta''''''''_{45}) / (\delta_5 {\otimes} \beta''''''''_{45})^{-D_5} =$

$\delta'_5(d_5,\ b_5,\ b_{55},\ b_3) =$

$(P(b_5| d_5, b_3) / P(b_{55}| b_5, b_3).\ \varphi_{2.300,\ 0.7672}\ (d_5))$

BIBLIOGRAPHY

[1] Afonja, B., (1972), "The moments of the maximum of correlated normal and *t*-variates," *Journal of the Royal Statistical Society*, Series B, **24**(2), 251–262.

[2] Anklesaria, K. P. and Z. Drezner (1986), "A multivariate approach to estimating the completion times for PERT networks," *Journal of the Operational Research Society*, **37**, 811–815.

[3] Breese, J. S., Heckerman, D., Kadie, C., "*Empirical Analysis of Predictive Algorithms for Collaborative Filtering*," In Proceedings of Fourteenth conference on Uncertainty in Artificial Intelligence, (1998)

[4] Bruno, D., Paché, G., "*From Traditional Retailing to E-tailing: The death and Rebirth of the Hypermarket Format*?" Frontiers of E-Business Research (2005)

[5] Burt, J. M., and M. B. Garman (1971), "Conditional Monte Carlo: A simulation technique for stochastic network analysis," *Management Science*, **18**, 207–217.

[6] Cannings,C., E. A. Thompson and M.H. Skolnick (1978),"Probability functions on complex pedigrees", *Advances in Applied Probability* , **10**, 26-61

[7] Cavoukian, A., "*Tag, You're It: Privacy Implications of Radio Frequency Identification (RFID) Technology*," Information and Privacy Commissioner (2004)

[8] Clark, C. E. (1961), "The greatest of a finite set of random variables," *Operations Research*, **9**(2), 145–162.

[9] Cobb, B. R. and P. P. Shenoy (2005), "Hybrid Bayesian networks with linear deterministic variables," in F. Bacchus and T. Jaakkola (eds.), *Uncertainty in*

Artificial Intelligence: Proceedings of the Twenty-First Conference (UAI-05), 136–144, AUAI Press, Corvallis, OR.

[10]   Cobb, B. R. and Shenoy, P.P (2006a)"Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials," International Journal of Approximate Reasoning, Vol. 41, No. 3, pp. 257--286,

[11]   Cobb, B. R. and P. P. Shenoy (2006b), "Operations for inference in continuous Bayesian networks with linear deterministic variables," *International Journal of Approximate Reasoning*, **42**(1–2), 21–36.

[12]   Cobb, B. R. and P. P. Shenoy (2007), "Propagation in hybrid Bayesian networks with linear deterministic variables," School of Business Working Paper No. 314, University of Kansas, Lawrence, KS.

[13]   Curtin, J., Kaufmann, R. J., Riggins, F. J., "*Making the Most Out of RFID Technology: A Research Agenda for the Study of Adaptation, Usage and Impact of RFID*", Information Technology and Management, (2007)

[14]   Decker, C., Kubach, U., Beigl, M., "*Revealing the Retail Black Box by Interaction Sensing*," In Proceedings of ICDCS 2003, Providence, Rhode Island (2003) 328–333.

[15]   Dipert, B., "*Reading between the Lines*," Business & Company Resource Center (2005)

[16]   Elmaghraby, S. E. (1967), "On the expected duration of PERT type networks," *Management Science*, **13**, 299–306.

[17]   Elmaghraby, S. E. (1977), *Activity Networks: Project Planning and Control by Network Models*, John Wiley and Sons, NY.

[18]   Finkenzeller, K., "RFID Handbook Radio-Frequency Identification and Applications," John Wiley & Son (1999)

[19]   Fishman, G. S. (1985), "Estimating critical path and arc probabilities in stochastic activity networks," *Naval Research Logistics Quarterly*, **32**, 249–261.

[20]   Fulkerson, D. R. (1962), "Expected critical path lengths in PERT networks," *Operations Research*, **10**, 808–817.

[21]   Garman, M. B. (1972), "More on conditional sampling in the simulation of stochastic networks," *Management Science*, **19**, 90–95.

[22]   Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R., Kadie, C., "*Dependency Networks for Inference, Collaborative Filtering, and Data Visualization*," Journal of Machine Learning Research 1 (2000) 49–75

[23]   Herlocker, J. L., Konstan, J. A., Riedl, J., "*Explaining Collaborative Filtering Recommendations*," Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work (2000) 241–250

[24]   Herlocker,J.L.Konstan, J.A., Terveen, L.G., Riedl, J., "*Evaluating Collaborative Filtering Recommender Systems*", ACM Trans on Information systems 22(1), (2004), 5-53

[25]   Hopwood, J., "*What is RFID*?" Intellident, http://www.intellident.co.uk

[26]   Inman, J., Winer, R., "W*here the Rubber Meets the Road: A Model for In-store Consumer Decision Making*," Marketing Science Institute Report, (1998) 98–122

[27]   Jenzarli, A. (1995), "Modeling dependence in project management," PhD dissertation, University of Kansas School of Business, Lawrence, KS.

[28]   Kelley, J. E. and M. R. Walker (1959), "Critical path planning and scheduling," in *Proceedings of the Eastern Joint Computer Conference*, pp. 160–173, Institute of Radio Engineers, NY.

[29]   Kenley, *C*.R. (1986), Influence diagram models with continuous variables, "PhD dissertation, Dept. of Engineering-Economic Systems, Stanford University

[30]   Kulkarni, V. G. and V. G. Adlakha (1986), "Markov and Markov-regenerative PERT networks," *Operations Research*, **34**(5), 769–781.

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

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

[33]   Lee, Y. M., Cheng, F., Leung, Y. T., "*Exploring the Impact of RFID on Supply Chain Dynamics*," Proceedings of the 2004 Winter Simulation Conference (2004)

[34]   Malcolm, D. G., J. H. Roseboom, C. E. Clark, and W. Fazar (1959), "Application of a technique for research and development program evaluation," *Operations Research*, **7**, 646–669.

[35]   Mild, A., "*An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data*," Adaptive Information Systems and Modelling in Economics and Management Science (2002)

[36]   Mital, T., "*The Emergence of RFID Technology,*" ISCR Future Technology Topic Brief, (2003)

[37]   Mittal, B., "*An Integrated Framework for Relating Diverse Consumer Characteristics to Supermarket Coupon Redemption,*" Journal of Marketing Research, Volume 31, November (1994) 533–544

[38]   Moral, S., Rumí, R., Salmeron, A. (2001) "Mixtures of truncated exponentials in hybrid Bayesian networks, Symbolic and Quantitative Approaches to Reasoning Under Uncertainty, Lecture Notes in Artificial Intelligence, Vol.2143, Springer Verlag, Heidelberg, pp. 156-167

[39]   Netflix, http://www.netflixprize.com/, (2007)

[40]   Nogués, M.T. .M., "Combining Machine Learning and Rule-Based Approaches in Spanish Syntatic Generation", PhD Dissertation, Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra (2006)

[41]   Olmsted, S.O. (1984), "On representing and solving decision problems," PhD dissertation, Dept. of Engineering-Economic Systems, Stanford University

[42]   Poland III, W. B. (1994), "Decision analysis with continuous and discrete variables: A mixture distribution approach" PhD dissertation, Dept. of Engineering–Economic Systems, Stanford University

[43]   Pazzani, M. J., "*A Framework for Collaborative, Content-Based and Demographic Filtering*," Artificial Intelligence Review (1999)

[44]   Robillard, P. (1976), "Expected completion time in PERT networks," *Operations Research*, **24**, 177–182.

[45]   Rumí, R. (2003) "Modelos De Redes Bayesianas Con Variables Discretas Y Continuas", Doctoral Thesis, Universidad de Almariá, Departamento de Estadística Y Matemática Aplicada, Almería, Spain

[46]   Schachter, R. D. (1986), "Evaluating influence diagrams," *Operations Research,* **34,** 871-882

[47]   Schachter, R. D. (1988), "*Probabilistic* inference and influence diagrams," *Operations Research*, **36**, 589-604

[48]   Schachter, R. D.(1990), "An ordered examination of influence diagrams," *Networks*, **20,** 535-563

[49]   Schachter, R. D. and C.R. Kenley (1989), "Gaussian influence diagrams," *Management Science*, **35**(5), 527-550

[50]   Sculli, D. (1983), "The completion time of PERT networks," *Journal of the Operational Research Society*, **34**(2), 155–158.

[51]   Shenoy, P. P. and G. Shafer (1990), "Axioms for Bayesian and belief function propagation," in R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 4*, 169–198, North-Holland, Amsterdam.

[52]   Shenoy, P.P (1992), "*Valuation-based systems: A framework for managing uncertainty in expert systems*", in Zadeh, L. A. and J. Kacprzyk(eds.), fuzzy Logic for the Management of Uncertainty, 83-104, John Wiley &Sons, New York, NY

[53]   Shenoy, P.P (1997), "Binary Join Trees for Computing Marginals in the Shenoy-Shafer Architecture", International Journal of Approximate Reasoning, **17**(2-3), 239-263

[54]   Shenoy, P. P. (2006), "Inference in hybrid Bayesian networks using mixtures of Gaussians," in R. Dechter and T. Richardson (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Second Conference* (UAI-06), 428–436, AUAI Press, Corvallis, OR.

[55]   Sigal, C. E, A. A. B. Pritsker, and J. J. Solberg (1979), The use of cutsets in Monte-Carlo analysis of stochastic networks," *Mathematics and Computers in Simulation*, **21**, 376–384.

[56]   Sinha, D., H. Zhou, and N. V. Shenoy (2006), "Advances in computation of the maximum of a set of random variables," *Proceedings of the 7th International Symposium on Quality Electronic Design* (ISQED-06), 306–311, IEEE Computer Society Washington, DC.

[57]   Van Slyke, R. M. (1963), "Monte Carlo methods and the PERT problem," *Operations Research*, **11**, 839–860.