

Modeling Correlated Discrete Uncertainties in Event Trees with Copulas

Tianyang Wang¹ James S. Dyer² John C. Butler³

Abstract

Modeling the dependence between uncertainties in decision and risk analyses is an important part of the problem structuring process. We focus on situations where correlated uncertainties are discrete, and extend the concept of the copula-based approach for modeling correlated continuous uncertainties to the representation of correlated discrete uncertainties. This approach reduces the required number of probability assessments significantly compared to approaches requiring direct estimates of conditional probabilities. It also allows the use of multiple dependence measures, including product moment correlation, rank order correlation and tail dependence, and parametric families of copulas such as normal copulas, t-copulas and Archimedean copulas. This approach can be extended to model the dependence between discrete and continuous uncertainties in the same event tree.

Keywords: Probabilistic risk analysis, decision analysis, event trees, discrete uncertainties, dependence, copulas

¹College of Business, Colorado State University, Fort Collins, Colorado 80523, Email: Tianyang.Wang@business.colostate.edu

²McCombs School of Business, University of Texas at Austin, Austin, Texas 78712, Email: Jim.Dyer@mcombs.utexas.edu

³McCombs School of Business, University of Texas at Austin, Austin, Texas 78712, Email: John.Butler2@mcombs.utexas.edu

Modeling Correlated Discrete Uncertainties in Event Trees with Copulas

1 Introduction

Probabilistic Risk Analysis (PRA) is a key methodology in the assessment of the likelihood and consequences of a system failure. Over the last thirty years the tool has been applied to the management of dams and water management systems (Wurbs 2005), the space shuttle (Paté-Cornell and Fischbeck 1994), nuclear facilities (EPRI 2004), healthcare (Cox 2007), nautical vessel management (Merrick et al. 2002), earthquakes (Budnitz et al. 1998) and terrorism (Ezell and Collins 2010). PRA is generally viewed as a mature discipline, but Bier and Cox (2007, pg. 280) note, “because of the complexity of the situations often addressed by PRA and the need for subjective judgments when inputs are uncertain, PRA still has . . . plenty of opportunity for methodological enhancements to improve the reliability, applicability, and acceptance of PRA results.” This paper addresses one such enhancement, modeling the dependence among discrete risks in a PRA.

Capturing the dependence among discrete risks would also be useful in a decision analysis setting where the choices of the decision maker are made explicit in a decision tree (e.g., Bickel and Smith 2006; Bickel, Smith and Meyer 2008). We will use the term “event tree” to refer to a collection of uncertainties represented by discrete event nodes that reflects the possible dependence among them. Our focus will be on the uncertainties in the analysis, but these correlated event trees could be featured in either a PRA or in a decision tree that also includes decision nodes.

Many uncertainties in PRA are represented as binary consequences because the events either occur or do not occur, or because that is a reasonable approximation of reality (e.g., Levenson 1995). For example the uncertain outcome of an oil and gas prospect may be

simplified as either successful in finding hydrocarbons, i.e. “wet”, or a dry hole (e.g., Bickel and Smith 2006). Discrete uncertainties with more than two branches are also common in decision and risk analyses. Some are naturally discrete, such as the number of times a terrorist could attack a target (zero, one, two or three). Other discrete events in probability trees are discrete approximations of continuous distributions. For example, the uncertain reserves of a successful oil well might be classified as low, medium, or high with those terms being associated with representative values from a continuous distribution, such as its 10th, 50th, and 90th percentiles. We will emphasize the distinction between naturally discrete probability distributions and discretized approximations to continuous distributions, as this distinction has implications regarding the appropriate approach to model dependencies among uncertainties.

The creation of an event tree is made much more challenging when there are dependencies among the events that should be modeled, and this is often the case. For example, the IAEA states (IAEA 2010, pg. 40) “dependent failures have often been found to be one of the dominant contributors to the [reactor] core damage frequency” in its discussion of the safety assessment of nuclear power plants. The significance of event dependence is exacerbated when discrete events are used to model the actions of an adversary in the context of two-player risk analysis and those actions are highly dependent upon precursor events (see discussions in Parnell et al. 2010 and Ezell and Collins 2010).

Event trees offer a natural way to show the dependence of one uncertainty on another in a graphical format. Typically, these dependence relationships are assessed by specifying the conditional distribution of each uncertainty given the discrete outcomes of the preceding uncertainty(s). Unfortunately, this approach is often impractical for subjective risk assessment, as the number of required conditional distributions grows exponentially with the number of dependent events (Keefer 2004). As a result, the risk analyst may ignore some or all of these dependencies in order to simplify the analysis (e.g. Stabell

2000, Delfiner 2003). As discussed in Keefer (2004), Abbas (2006) and Bickel and Smith (2006), this type of simplification may cause significant errors.

In this paper we extend the copula-based approach (Clemen and Reilly 1999, Wang and Dyer 2012) for modeling correlated continuous uncertainties to the representation of correlated discrete uncertainties. In both of these approaches, the analyst must obtain the parameters of continuous probability distributions from historical data or from assessments from experts and obtain a measure of the dependence between them in the form of an estimate of correlation as discussed by Clemen, Fisher, and Winkler (2000). For example, one of two correlated continuous distributions might be a beta distribution with estimated parameter values α and β and the other might be a normal distribution with estimated values for the mean and the standard deviation. If the normal distribution is to be conditioned on the beta distribution, then a three outcome discrete event approximation to the beta would be made using one of the approximately moment matching approaches such as the extended Pearson-Tukey (EPT) method (Keefer 1983). While there are many approaches to discretize continuous distributions (see Hammond and Bickel 2013), each of them would be applicable with the proposed methodology. For convenience, we will assume the analyst has chosen the EPT discretization method in our illustrative examples.

Then the continuous copula-based approaches would estimate three continuous normal distributions conditioned on the three outcomes of the discrete approximation to the beta distribution using the measure of correlation between them. In turn, each of these three conditional distributions would be discretized using the EPT method so that they would also be approximately moment matching discrete distributions. In summary, these three conditional discrete event nodes are approximations of the conditional continuous distributions obtained from one of the original continuous marginal distributions and a measure of its dependence on the other continuous distribution.

In this paper, we show how to construct the dependent event tree when uncertainties

are discrete, as a complement to the Wang and Dyer (2012) approach, hereafter W&D. The approach for discrete uncertainties has to be modified because W&D requires that all marginal distributions are continuous and differentiable. We also extend both approaches so that copulas can be used to capture dependence when there are both continuous and discrete uncertainties in the same event tree.

A natural question that arises is the following. Since the W&D approach uses a discretization of the first uncertainty and calculates continuous conditional distributions that are then discretized, why not simply discretize both distributions initially using the EPT and use this new methodology on these two discrete approximations? As we shall discuss, the results of this latter approach would be different from the results using the W&D approach for the continuous distributions, and the implications of this observation offer some insights regarding the methods for discretizing continuous distributions that have heretofore been accepted as standard practice in many decision and risk analysis applications.

The remainder of this paper is organized as follows. We begin by assuming that we have discrete uncertainties and derive a copula approach for modeling their dependence. Section 2 provides a closed-form solution for the discrete dependence relationship between two binary uncertainties when product moment correlations are provided, and contrasts the copula-based approach with other methods when the closed-form solution cannot be obtained. Section 3 develops the analytical form of the joint and conditional probabilities for binary uncertainties based on copulas when product moment correlations or rank order correlations are provided and compares our example to other methods while Section 4 extends our approach to n-branch uncertainties. In section 5 we develop a copula approach to capture dependencies when there are both natural continuous and discrete uncertainties in the same event tree and highlight the differences between this new approach for discrete uncertainties and W&D. Section 6 compares the distinction between discretizing

a naturally continuous uncertainty and then using this discrete copula approach and applying the W&D approach to the continuous uncertainties directly. Section 7 provides a summary and conclusion.

2 Estimation of Discrete Conditional Probabilities Using Correlations

We begin with an analysis of a pair of binary uncertainties to serve as a basis of comparison with more complex cases, and derive a closed-form solution for the probabilities of one event conditioned on another. However, a similar closed-form solution for the dependence relationships does not exist when there are more than two binary uncertainties or when discrete uncertainties have more than two outcomes.

2.1 Conditional Probabilities for Correlated Binary Uncertainties

Suppose we have two dependent binary uncertainties representing whether or not the events X_1 and X_2 occur with marginal probabilities p_1 and p_2 respectively. For simplicity, we refer to both the uncertainties and the related events as X_1 and X_2 in this section. We use 1 and 0 as indicator outcomes for these events, so that p_1 is equal to the probability that the outcome associated with event $X_1 = 1$, and p_2 is the probability that the outcome associated with event $X_2 = 1$. Similarly p_{12} is the joint probability that $X_1 = 1$ and $X_2 = 1$, and $p_{2|X_1=k}$ is the conditional probability that $X_2 = 1$ given $X_1 = k, k \in (0, 1)$. The product moment correlation r_{12} is defined in terms of the dependence between these

indicator outcomes, and is independent of any consequences associated with them. This is equivalent to the representation of discrete binary events used by Bickel and Smith (2006).

Using the definition of the marginal probability of X_2 , p_2 , and the product moment correlation between these binary uncertainties, r_{12} , we have

$$p_2 = p_1 p_{2|X_1=1} + (1 - p_1) p_{2|X_1=0} \text{ and } r_{12} = \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{E(X_2|X_1)E(X_1) - E(X_1)E(X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{p_{2|X_1=1}p_1 - p_1 p_2}{\sqrt{p_1 - p_1^2} \sqrt{p_2 - p_2^2}}.$$

Given these two equations in two unknowns, the formulas for the conditional probabilities of X_2 given X_1 can be derived as

$$p_{2|X_1=1} = \sqrt{\frac{1-p_1}{p_1}} \sqrt{p_2 - p_2^2} r_{12} + p_2, \text{ and } p_{2|X_1=0} = p_2 - \sqrt{\frac{p_1}{1-p_1}} \sqrt{p_2 - p_2^2} r_{12}.$$

Bickel, Smith and Meyer (2008) provide the same relationship in a different form.

We now consider a third binary uncertainty representing whether or not the event X_3 occurs that is also dependent on the binary uncertainties X_1 and X_2 . There are four conditional probability relationships that include X_3 which need to be specified, $p_{3|X_1=0, X_2=1}$, $p_{3|X_1=1, X_2=0}$, $p_{3|X_1=0, X_2=0}$ and $p_{3|X_1=1, X_2=1}$. However, with the given marginal probability p_3 and correlations r_{13} and r_{23} , we have only the following three equations relating the conditional probabilities of event X_3 to events X_1 and X_2 :

$$\begin{aligned} p_{3|X_1=0, X_2=1} &= \frac{r_{23} \sqrt{p_2 - p_2^2} \sqrt{p_3 - p_3^2} + p_2 p_3 - p_{3|X_1=1, X_2=1} p_2 |X_1=1 p_1}{p_{2|X_1=0} (1 - p_1)}, \\ p_{3|X_1=1, X_2=0} &= \frac{r_{13} \sqrt{p_1 - p_1^2} \sqrt{p_3 - p_3^2} + p_1 p_3 - p_{3|X_1=1, X_2=1} p_2 |X_1=1 p_1}{(1 - p_{2|X_1=1}) p_1}, \text{ and} \\ p_{3|X_1=0, X_2=0} &= \frac{p_3 (1 - p_1 - p_2) - r_{13} \sqrt{p_1 - p_1^2} \sqrt{p_3 - p_3^2} - r_{23} \sqrt{p_2 - p_2^2} \sqrt{p_3 - p_3^2} + p_{3|X_1=1, X_2=1} p_2 |X_1=1 p_1}{(1 - p_{2|X_1=0}) (1 - p_1)}. \end{aligned}$$

We lack a relationship for $p_{3|X_1=1, X_2=1}$ in the underlying dependence structure and therefore cannot find a unique solution for these conditional probabilities. We refer to this situation as one that is under-specified.

In general, when there are n uncertainties, only n marginal probabilities and $\frac{n(n-1)}{2}$

pairwise correlations can be specified. The number of conditional probabilities required to define the joint binary distribution grows exponentially, but the correlation information only grows combinatorially. This leads to situations with fewer equations than unknowns, so the required conditional probabilities cannot be uniquely determined. In some special cases with equal marginal probabilities or with an exchangeable correlation matrix where pairwise correlations are all the same, the joint distribution is uniquely specified (Lunn and Davies 1998, Oman and Zucker 2001, Papathomas 2008). However, for a binary event tree with more than two uncertainties, there is no *general* unique solution for the conditional probabilities as a function of the marginal probabilities and the pairwise product moment correlations.

2.2 Approaches to Deal with Under-specification

The problem of under-specification has long been recognized in the literature (e.g., Sarin 1979; Abbas 2006; Bickel and Smith 2006). Numerous approximation methods have been developed to specify a unique joint probability distribution given the available partial information.

Keefer (2004) introduced the underlying event model, which requires the assessment of only one conditional probability in addition to the marginal probabilities, and determines the entire joint probability distribution assuming that all random variables are conditionally independent given another “hidden” or “underlying” random variable. The potential error associated with using the underlying event model has been discussed by Abbas (2006) and Bickel and Smith (2006) because of its assumptions of independence among most binary events and its ability to accommodate only positive dependence.

Jaynes (1957) first developed the principle of maximum entropy (ME) to avoid the need to elicit the complete joint probability of uncertainties. For example, adopting the

notation of Abbas (2006) for maximum entropy formulation, when there are three binary uncertainties X_1 , X_2 and X_3 , maximum entropy is defined as $p_{1,2,3}^* = \underset{p_{1,2,3}}{argmax} - \sum_{1,2,3} p_{1,2,3} \log(p_{1,2,3})$, where $p_{1,2,3} = p_{3|X_1=1, X_2=1}$. The ME method specifies the joint distribution that maximizes entropy and is closest to the uniform distribution, subject to information that is available. An equivalent approach is to choose the joint distribution that minimizes the Kullback-Leibler divergence to a uniform distribution that assumes independence (Kullback and Leibler 1951).

ME has been proposed and used in modeling uncertainties in decision analysis (e.g. Abbas, 2006; Bickel and Smith, 2006; Bickel, Smith and Meyer, 2008). Abbas (2006) illustrated the use of the ME method to obtain a joint probability distribution using lower order joint probability assessments in the context of a semiconductor testing application. He showed that errors due to the inappropriate assumption of event independence can lead to suboptimal decisions. Bickel and Smith (2006) developed a practical ME framework for evaluating sequential exploration strategies in oil exploration. In the case where the exploration prospects are dependent, they applied the ME method to construct a full joint distribution for all outcomes from the marginal and pairwise assessments. Based on the constructed discrete joint distributions and a dynamic programming model, their methodology can determine the optimal exploration strategies.

More recently, Montiel and Bickel (2013) augmented the ME approximation with the analytic center (AC) method which maximizes the product of the joint probabilities instead of maximizing the entropy in the ME method, and also proposed using the average of all simulated joint distributions that are consistent with the lower order assessments. They show that these methods produce results similar to the ME method by using the oil exploration example introduced by Bickel and Smith (2006). In practice, these methods require customized coding and the determination of when it is appropriate to assume that the average of a large number of simulated candidate joint distributions is representative of

the true joint distribution; these issues are exacerbated when the number of uncertainties is large.

Like the ME approach, copulas can also be used to provide the additional information that avoids under specification and yields a unique joint distribution given marginal probabilities and dependence information (c.f. Jouini and Clemen 1996; Embrechts, et al., 1999; Cherubini et al., 2004). A copula allows a joint distribution of random variables to be expressed as a function of the marginal distributions. For a random vector $X = (X_1, \dots, X_n)$ with $F_i(X_i)$ denoting the marginal distribution of each X_i , Sklar (1959) showed that there always exists a function C such that $F(X_1, \dots, X_n) = C(F_1(X_1), \dots, F_n(X_n))$. The function C is called a copula. It fully captures the dependence structure among the random variables, and is independent from the choice of the marginal distributions.

The copula function C is itself a distribution function for uniform random variables since the marginal CDFs are standard uniform distributions. Let $U_1 = F_1(X_1)$, ..., $U_n = F_n(X_n)$, the joint CDF can be given by the copula $C(U_1 = u_1, \dots, U_n = u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$, where F^{-1} denotes the inverse cumulative distribution function, and $U_i, i = 1, \dots, n$, is a uniform variable defined over $[0,1]$.

The use of a copula allows the independent consideration of the marginal probabilities and the dependence structure. The joint distribution of $X = (X_1, \dots, X_n)$ can be reconstructed from the copula and the marginal distributions $F_i(X_i), i = 1, \dots, n$. Using just these marginal distributions and information about the dependence structure, it is possible to model continuous dependent multivariate uncertainties in an event tree (Clemen and Reilly 1999, Wang and Dyer 2012). Here we focus on capturing the correlation among discrete uncertainties as opposed to continuous uncertainties.

The separation of the marginal probabilities and the dependence between those probabilities in a copula suggests an effective algorithm for constructing event trees: (1) determine the marginal probability distributions; (2) estimate the correlations among these

probability distributions; (3) combine the dependence information with the marginal probability distributions to calculate the desired conditional probabilities in the constructed binary event tree.

In the following sections we apply our approach to a variety of settings and discuss implementation issues. We will make specific assumptions and about the measures of dependence and the appropriate copula family, and in practice these choices are important issues. W&D provide a detailed discussion of a variety of dependence measures and copula families that can be applied to generate correlated continuous event trees and that are equally applicable for discrete event trees. For more information about assessing correlation readers are referred to Clemen, Fisher, and Winkler (2000).

3 Modeling Dependence among Discrete Uncertainties using Copulas

In this section we start with a demonstration of the use of copulas to capture dependence between a pair of binary uncertainties. We then extend the analysis to situations with more than two binary uncertainties and compare our approach to ME using an example from Bickel and Smith (2006).

3.1 The Case of Binary Uncertainties

The copula function for a pair of dependent binary uncertainties is

$$F(X_1, X_2) = C(F_1(X_1), F_2(X_2)) = C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))$$

As before, we define p_i such that $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$ where p_i is the $100p_i^{th}$ percentile of the uniform distribution and r_{ij} is the product moment correlation between uncertainties X_i and X_j . For instance, the joint distribution

of both $X_1 = 1$ and $X_2 = 1$ can be represented by a copula:

$$F(X_1 = 1, X_2 = 1) = C(p_1, p_2) = \int_0^{p_1} \int_0^{p_2} c(u_1, u_2) du_1 du_2.$$

When the dependence structure between the random variables is captured by a bivariate normal copula, C is given by $C(F_1(X_1), F_2(X_2)) = \Phi_{\rho_{12}}(\Phi^{-1}(F_1(X_1)), \Phi^{-1}(F_2(X_2)))$, where Φ is the cumulative distribution function for a standard bivariate normal distribution function with mean zero and correlation ρ_{12} and the marginal cumulative distributions of $F_1(X_1)$ and $F_2(X_2)$ are transformed by the inverse of the standard normal distribution CDF function Φ^{-1} .

The normal copula shares the same dependence structure as the multivariate normal distribution; it uses pair-wise Pearson product moment correlations to measure dependence and allows arbitrary marginal distributions for the uncertainties. Because this copula uniquely determines the dependence structure, there is a one-to-one relationship between the product moment correlation ρ_{12} for the normal copula function and the product moment correlation r_{12} between the original uncertainties. Cairo and Nelson (1997) provide efficient algorithms to determine the normal copula correlations through a correlation matching technique when Pearson product moment correlations are used to measure dependence. For example, the correlation matching formulation of bivariate normal copulas satisfies the following equation:

$$r_{12} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2,$$

where ϕ is the density function for a standard bivariate normal distribution function with mean zero and correlation ρ_{12} , and z_1 and z_2 are standard normal variables.

With bivariate normal copulas, the joint probability of $X_1 = 1, X_2 = 1$, can be defined as

$$F(X_1 = 1, X_2 = 1) = \int_{-\infty}^{\Phi^{-1}(p_1)} \int_{-\infty}^{\Phi^{-1}(p_2)} \phi(z_1, z_2, \rho_{12}) dz_1 dz_2.$$

The other joint probabilities can be calculated using the relationships between the marginals or using copulas. For instance, $F(X_1 = 1, X_2 = 0) = p_2 - F(X_1 = 1, X_2 = 1)$, or alternatively,

$$F(X_1 = 1, X_2 = 0) = \int_{-\infty}^{\Phi^{-1}(p_1)} \int_{\Phi^{-1}(p_2)}^{\infty} \phi(z_1, z_2, \rho_{12}) dz_1 dz_2.$$

The conditional probabilities may be written

$$F(X_2 = j | X_1 = i) = \frac{F(X_1=i, X_2=j)}{F(X_1=i)}, i, j = 0, 1; \text{ e.g.,}$$

$$p_{2|X_1=1} = \frac{F(X_1 = 1, X_2 = 1)}{p_1} = \frac{\int_{-\infty}^{\Phi^{-1}(p_1)} \int_{-\infty}^{\Phi^{-1}(p_2)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{p_1} \quad (1)$$

This equation differs from the development in W&D. As we will discuss in section 5.1, the W&D method works only for continuous distributions. Intuitively, W&D calculates the conditional percentiles of the subsequent continuous uncertainty while keeping the conditional probabilities constant consistent with the EPT method. When the uncertainties are discrete, it is meaningless to calculate the conditional percentiles of the discrete uncertainties because all possible discrete outcomes are already predetermined (e.g. head and tail of the flip coins). Therefore, in the proposed method uniquely designed for the discrete uncertainties, we calculate the conditional probabilities while keeping the discrete outcomes as given.

The copula family utilized in the analysis implies a unique dependence structure for the underlying binary uncertainties and hence a unique event tree structure. From our previous derivation, we know that there is unique solution for the bivariate tree with only two uncertainties. Therefore, if the normal copula is the underlying copula for this dependence structure, the joint probabilities calculated from the normal copulas must agree with the formulas derived directly from the marginal probabilities and the product moment correlation.

Substituting (1) into $r_{12} = \frac{p_{2|X_1=1}p_1 - p_1p_2}{\sqrt{p_1 - p_1^2}\sqrt{p_2 - p_2^2}}$ and assuming a normal copula we can derive

the following relationship

$$r_{12} = \frac{\int_{-\infty}^{\Phi^{-1}(p_1)} \int_{-\infty}^{\Phi^{-1}(p_2)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2 - p_1 p_2}{\sqrt{p_1 - p_1^2} \sqrt{p_2 - p_2^2}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2.$$

which can be verified numerically for $\rho_{12} \in [-1, 1]$. This is an interesting result as it shows that for *any* two binary uncertainties, the underlying dependence structure is defined by a bivariate normal copula; note that we did not assume a normal copula when deriving the relationship. For a pair of binary uncertainties, the ME and AC methods give the same result obtained from using the normal copula and these formulas. This is consistent with previous work demonstrating that the use of a Normal copula leads to a “near-maximum entropy” result in general (Bethge and Berens 2008, Macke et al. 2009).

In the case of n binary uncertainties, the copulas-based joint and conditional probabilities are given by

$$F(X_1 = i_1, \dots, X_n = i_n) = \int_{lb_1}^{ub_1} \dots \int_{lb_n}^{ub_n} c(u_1, \dots, u_n) du_1 \dots du_n, \text{ and}$$

$$F(X_n = i_n | X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = \frac{F(X_1=i_1, \dots, X_n=i_n)}{F(X_1=i_1, \dots, X_{n-1}=i_{n-1})}, \text{ respectively,}$$

where $i_j = 0, 1, j = 1, \dots, n$, and $lb_i = \begin{cases} 0 & X_i = 1 \\ p_i & X_i = 0 \end{cases}$, $ub_i = \begin{cases} p_i & X_i = 1 \\ 1 & X_i = 0 \end{cases}$.

With normal copulas defining the underlying dependence structure, the joint probability is

$$F(X_1 = i_1, \dots, X_n = i_n) = \int_{lb_1}^{ub_1} \dots \int_{lb_n}^{ub_n} \phi(z_1 \dots z_n, \Sigma_Z) dz_1 \dots dz_n,$$

where $i_j = 0, 1, j = 1, \dots, n$, $lb_i = \begin{cases} \Phi^{-1}(p_i) & X_i = 0 \\ -\infty & X_i = 1 \end{cases}$, $ub_i = \begin{cases} \infty & X_i = 0 \\ \Phi^{-1}(p_i) & X_i = 1 \end{cases}$.

3.2 A Numerical Example with Binary Uncertainties

As previously discussed, with a pair of binary uncertainties there is no under specification as there is an exact equation relating the correlation and the joint and conditional probabilities. We will illustrate the case of three binary uncertainties using a subset of the event tree in Bickel and Smith (2006). In their example there are six sites where an oil well could be drilled, and each has a unique probability of success, which is defined as being “wet”. We focus on the first three drilling opportunities to provide a visual

comparison of the conditional probabilities generated by each approach, but the example could easily be expanded to include all six.

Figure 1 shows the marginal probability of well i being wet, p_i , and the correlations between each of the first three wells from this example. Using this data and the spreadsheet provided by the authors, the resulting ME event tree is shown in Figure 2. The positive correlation between the wells is captured by the conditional probabilities calculated by the ME algorithm. For example, the conditional probability $p_{3|X_1=1, X_2=1} = 0.717$ is greater than the marginal probability $p_3 = 0.53$ for the success of well 3, and the joint probability (shown in bold at end of each branch) of all three wells being “wet” is 0.148 compared to 0.091 if all three wells were independent.

Figure 1: Marginal Probabilities and Pearson Correlation (r) for First Three Wells in Bickel & Smith (2006)

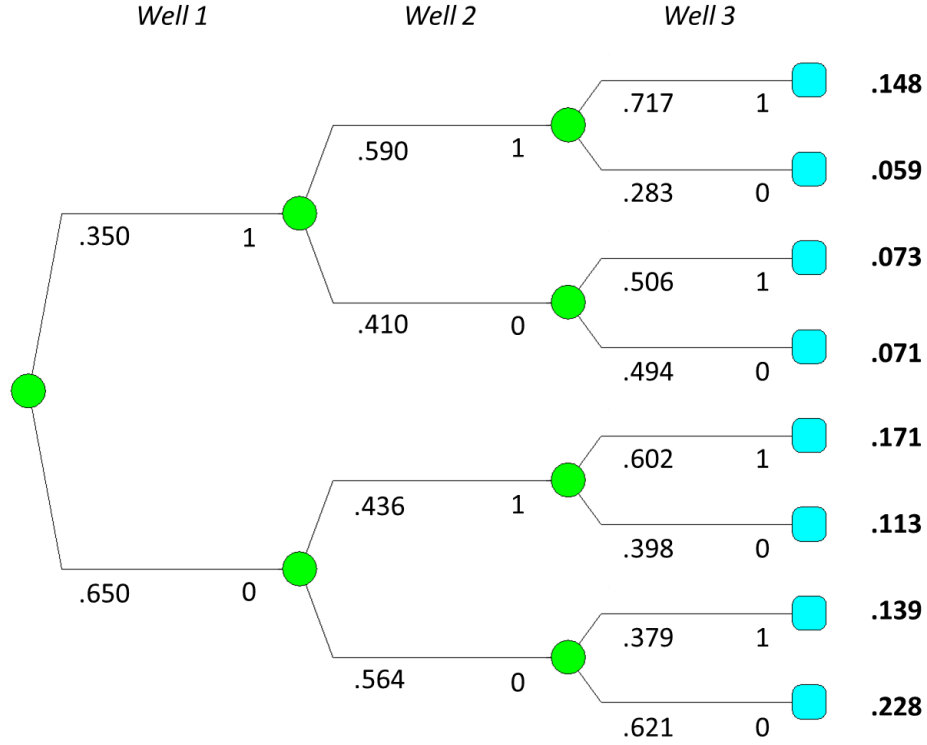
	P(Wet)	Well i / j	1	2	3
Well 1	0.35	1	1.0000	0.1468	0.1470
Well 2	0.49	2		1.0000	0.2357
Well 3	0.53	3			1.0000
Marginal Probabilities		Correlation Matrix			

The application of the copula approach with a normal copula described in the previous section resulted in the event tree shown in Figure 3. For example, the conditional probability $p_{2|X_1=1}$ is calculated as

$$\begin{aligned}
p_{2|X_1=1} &= \frac{F(X_1=1, X_2=1)}{F(X_1=1)} = \frac{\int_{-\infty}^{\Phi^{-1}(p_1)} \int_{-\infty}^{\Phi^{-1}(p_2)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{p_1} \\
&= \frac{\int_{-\infty}^{\Phi^{-1}(0.35)} \int_{-\infty}^{\Phi^{-1}(0.49)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{0.35} = 0.590.
\end{aligned}$$

Consistent with the observations of Bethge and Berens (2008) and Macke et al. (2009) that this would be a “near maximum entropy” solution, the results rounded to three decimal places are almost identical; the only difference is $p_{3|X_1=1, X_2=1} = 0.716$ with the Normal copula versus 0.717 with ME. Thus the copula approach provides a closed-form

Figure 2: Event Tree for the Three Well Example using ME

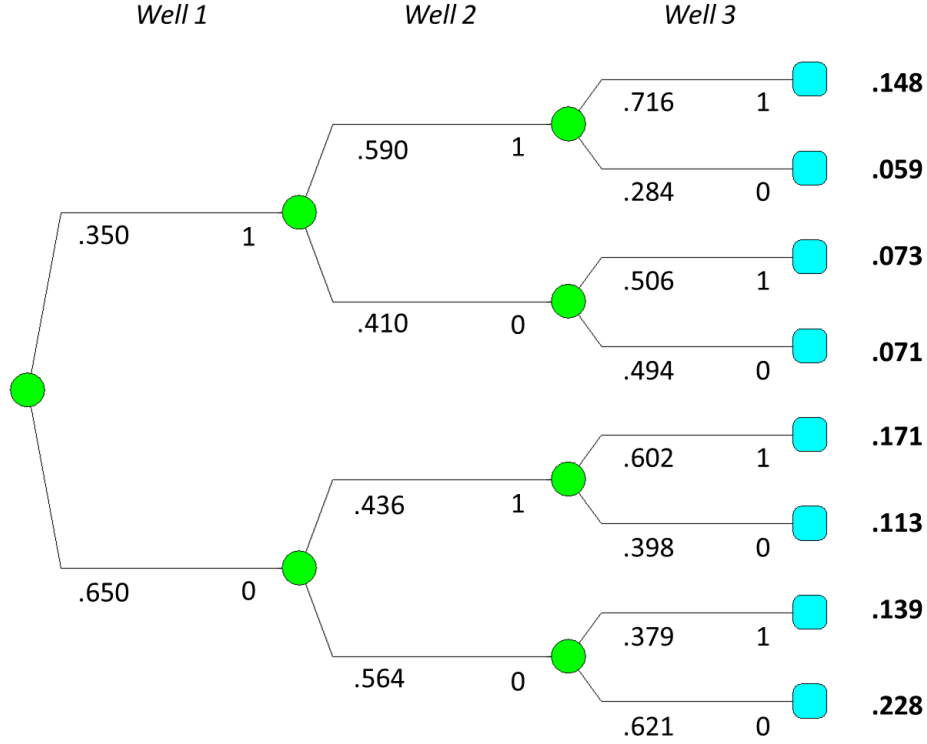


solution that closely approximates the ME solution. We also constructed the complete tree for the 6 wells example in Bickel and Smith (2006) and confirmed the “near maximum entropy” solution of the normal copulas-based tree (entropy = 3.331) compared to the optimization-based tree by Bickel and Smith (2006) (entropy= 3.332).

3.3 Comments on the Copulas-based Approach and the Optimization-based Approaches

As illustrated in Figures 2 and 3, the ME and copula approach offer similar solutions to the same problem; while we do not demonstrate it here, the AC approach could also be used. Further, the copula and optimization-based approaches can address the exponential

Figure 3: Event Tree the Three Well Example using the Normal Copula and the Pearson Correlation (r)



growth in the required number of assessments. For example, Bickel and Smith (2006) assess all pairwise conditional probabilities and show how one could achieve the same level of information about the joint distribution if one assessed the same number of pairwise correlations.

The proposed copulas-based approach does offer some advantages over the ME and AC approaches. First, the copulas-based approach provides a closed-form solution for the joint and conditional probabilities that can be solved numerically. In comparison, the calculation of the maximum entropy distribution of correlated variables with pre-specified marginal distributions requires the solution of nonlinear coupled integral equations subject to local optima, with no exact solution for discrete marginal distributions (Larralde 2012). Second, this approach allows relatively easy incorporation of nonlinear measures

of association, such as rank order correlation and tail dependence, in the same general framework, while the ME and AC will require customized coding for each application (for example, see Montiel and Bickel (2012) for a discussion of the constraints for rank correlations). Third, as demonstrated above the use of the multivariate normal copula provides a computationally more efficient approximation to the ME method that shares the benefits of a “near-maximum entropy” result while reducing its practical limitations (Bethge and Berens 2008; Macke et al. 2009). In addition, as we discuss in Section 5, the proposed approach allows the mix of dependent continuous and discrete uncertainties in the constructed event tree, which has not been discussed in the previous literature.

4 The Case of Discrete Uncertainties with More than Two Events

As discussed in sections 2 and 3, while a closed form solution relating dependence to marginal and conditional probabilities exists for a pair of binary uncertainties, when there are more than two uncertainties we encounter under-specification. The same is true for a pair of uncertainties when at least one has more than two events.

4.1 Extending the Copula-based Method to Uncertainties with More than Two Events

The copulas-based approach can be extended to the general case of dependent uncertainties with more than two discrete events as we illustrate with a simple example of two uncertainties X_i , $i = 1, 2$ with discrete probabilities p_i^u , p_i^m , p_i^d of three outcomes, $u =$ up, $m =$ middle, $d =$ down and the Pearson product moment correlation r_{12} . We define the 3-outcome indicator variable $X_i = 1$ with probability p_i^u ; $X_i = 0$ with probability p_i^m ,

and $X_i = -1$ with probability p_i^d . Note that the choice of 1, 0, -1 as indicator variable outcomes is an arbitrary one as long as these outcomes are equally spaced, e.g. 1, 0, -1. The results will be the same for any positive affine transformation of the indicator variables due to the relationship $r_{aX+b, cY+d} = \frac{acCOV(X,Y)}{a\sigma_X c\sigma_Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = r_{XY}$ for the product moment correlation.

If we define $X_i = -1$ if $u_i \leq p_i^d$, $X_i = 0$ if $p_i^d < u_i \leq p_i^d + p_i^m$ and $X_i = 1$ otherwise, where p_i is the $100p_i^{th}$ percentile of the uniform distribution, the joint distribution can be defined by copulas. For n events, the copulas-based joint and conditional probabilities are given by

$$F(X_1 = i_1, \dots, X_n = i_n) = \int_{lb_1}^{ub_1} \dots \int_{lb_n}^{ub_n} c(u_1, \dots, u_n) du_1 \dots du_n, \text{ and}$$

$$F(X_n = i_n | X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = \frac{F(X_1=i_1, \dots, X_n=i_n)}{F(X_1=i_1, \dots, X_{n-1}=i_{n-1})}, \text{ respectively,}$$

$$\text{where } i_j = -1, 0, 1, j = 1, \dots, n, \text{ and } lb_i = \begin{cases} p_i^d + p_i^m & X_i = 1 \\ p_i^d & X_i = 0 \\ 0 & X_i = -1 \end{cases}, ub_i = \begin{cases} 1 & X_i = 1 \\ p_i^d + p_i^m & X_i = 0 \\ p_i^d & X_i = -1 \end{cases}.$$

If we apply normal copulas to model the underlying dependence structure, then

$$F(X_1 = i_1, \dots, X_n = i_n) = \int_{lb_1}^{ub_1} \dots \int_{lb_n}^{ub_n} \phi(z_1, \dots, z_n, \Sigma_Z) dz_1 \dots dz_n, \text{ where } i_j = -1, 0, 1, j = 1, \dots, n,$$

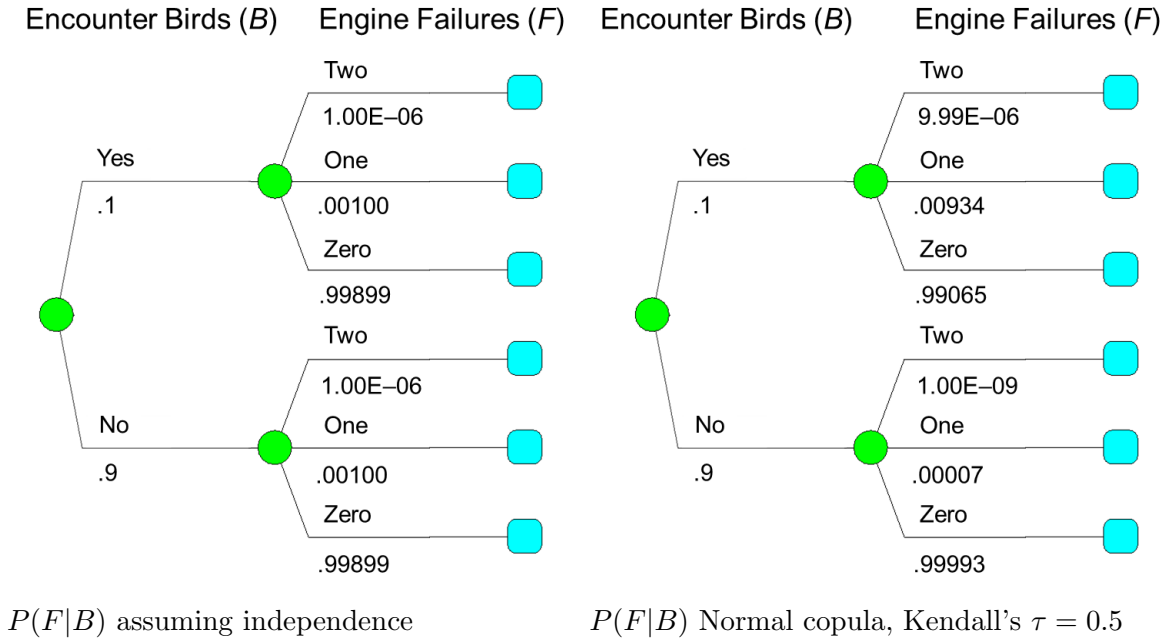
$$\text{and } lb_i = \begin{cases} \Phi^{-1}(p_i^d + p_i^m) & X_i = 1 \\ \Phi^{-1}(p_i^d) & X_i = 0 \\ -\infty & X_i = -1 \end{cases}, ub_i = \begin{cases} \infty & X_i = 1 \\ \Phi^{-1}(p_i^d + p_i^m) & X_i = 0 \\ \Phi^{-1}(p_i^d) & X_i = -1 \end{cases}.$$

To build the multivariate copula-based event tree for discrete (X_1, \dots, X_n) , we begin with X_1 and then recursively compute the conditional probabilities for the discrete uncertainties $X_k (k = 2, \dots, n)$, conditioning on each of the indicator variable values of the previous realizations for (X_1, \dots, X_{k-1}) . Throughout the tree, the indicator variable values of these approximations do not change, but the conditional probabilities assigned for the branches of a chance node vary with the conditioning scenarios using the logic of the copulas approximation. The fact that the probabilities, not the indicator values, vary to capture the correlation will be explored again in Section 6.

4.2 A Numerical Example of Discrete Uncertainties with More than Two Events

We will illustrate the use of modeling dependence between two discrete events with a hypothetical example regarding two uncertainties in the risk analysis of a twin engine plane: encountering birds on takeoff (B) and engine failure (F). As shown in Figure 4, we either do or do not encounter birds and the aircraft is assumed to have only two engines so we either have zero, one, or two engine failures on takeoff. The event tree on the left side of Figure 4 is based on the assumption that B and F are independent and therefore shows the hypothetical marginal distributions of each uncertainty. For the sake of clarity we used an artificially high probability of one or more engine failures but the approach would apply for any marginal distribution.

Figure 4: Probability of Engine Failure given Encounter with Birds



Encountering birds at takeoff increases the risk of engine failure, so we would expect $P(F = Two|B = Yes) > P(F = Two|B = No)$. The conditional probabilities of engine

failure on the right side of Figure 4 were calculated based on the assumption that the measures of dependence are provided by a Kendall's tau with $\tau = 0.5$ and a normal copula. For example, the conditional probability $p_{F=One|B=Yes}$ is calculated as

$$\begin{aligned} p_{F=One|B=Yes} &= \frac{F(F=One, B=Yes)}{F(B=Yes)} = \frac{\int_{-\infty}^{\Phi^{-1}(p_1)} \int_{\Phi^{-1}(p_2^d)}^{\Phi^{-1}(p_2^d + p_2^m)} z_1 z_2 \phi(z_1, z_2, \rho) dz_1 dz_2}{p_1} \\ &= \frac{\int_{-\infty}^{\Phi^{-1}(0.1)} \int_{\Phi^{-1}(0.000001)}^{\Phi^{-1}(0.000001 + 0.001)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{0.1} = 0.00934. \end{aligned}$$

Note that the corresponding conditional probabilities of one or two engine failures given that birds are encountered (Yes) are higher than their marginal values, and even higher than the corresponding conditional probabilities if birds are not encountered (No).

Many probabilistic risk analyses are conducted to find paths of high probability or high consequence or both. Comparing the left and right sides of Figure 4, $P(F = Two|B = Yes)$ is about $(0.1 \times 9.99E - 06) / (0.1 \times 1.00E - 06) = 9.9$ times higher when we explicitly capture the assumed dependence of Engine Failure on Birds Encountered. Assuming that B and F are independent if they are not could lead to errors in estimating the risk of air travel under different scenarios based on this hypothetical example.

5 Event Trees with both Continuous and Discrete Uncertainties

In reality many decision and risk analyses involve a mix of discrete and continuous uncertainties, and we can extend the proposed approach to handle these situations. While we illustrate the combination of discrete and continuous uncertainties with the simplest case of one discrete and one continuous uncertainty, the process can be generalized to multiple discrete and continuous uncertainties in any sequence. The discussion in this section

also allows us to contrast the technical aspects of the proposed discrete approach with the continuous W&D approach; in section 6, we demonstrate the practical distinctions between the two.

5.1 Extending the Copula-based Approach for Mixed Uncertainties

For the case of two conditional continuous uncertainties that were the focus of W&D, for each given conditional percentile $Pr(X_2 \leq x_2 | X_1 = x_1) = \alpha_2$, Darsow et al. (1992) show that $Pr(X_2 \leq x_2 | X_1 = x_1) = \frac{\partial C}{\partial F_1(x_1)}(F_1(x_1), F_2(x_2)) = \frac{\partial C}{\partial u_1}(u_1, u_2)$. Hence, the dependent uniform variable $u_2 = Pr(X_2 \leq x_2)$ is the inverse function of the realization of $u_1 = \alpha_1$ and the choice of the percentile α_2 for the conditional distribution $X_2 \leq x_2 | X_1 = x_1$. Let $\frac{\partial C}{\partial u_1}(u_1, u_2) = c_{u_1}(u_2) = \alpha_2$, then $u_2 = c_{u_1}^{-1}(\alpha_2)$. As illustrated in W&D, with a given copula we know the partial derivative and can derive the closed-form formula for u_2 . When the underlying copula is normal, $\frac{\partial C}{\partial u_1}(u_1, u_2) = \Phi\left(\frac{\Phi^{-1}(u_2) - r\Phi^{-1}(u_1)}{\sqrt{1-r^2}}\right)$ and $u_2 = \Phi(r\Phi^{-1}(\alpha_1) + \sqrt{1-r^2}\Phi^{-1}(\alpha_2))$. When we attempt to model discrete uncertainties, we have to modify the approach taken to find the conditional percentiles which represents the primary extension to W&D.

The combined methodology depends on the sequence of the uncertainties: continuous before discrete or vice versa. When a discrete uncertainty is preceded by a continuous uncertainty, we use the discretized continuous uncertainties to calculate the parameters to be used by the discrete uncertainty. When a continuous uncertainty is preceded by a discrete uncertainty, we use the discrete version of the copulas-based conditional percentiles to calculate the parameters to be used by the subsequent continuous uncertainty. We begin with the situation that is visually the most similar to W&D in an event tree, a discrete distribution conditioned on a continuous distribution and then move on to the reverse order.

When a discrete event follows a continuous uncertainty in the sequence we can use the W&D approach for the continuous variables and then numerically calculate the integral bounds for the subsequent discrete uncertainty. First, we build the discrete approximation for the preceding continuous uncertainty X_1 with the 5th, 50th and 95th percentiles according to the EPT method, which corresponds to the realizations of 0.05, 0.5, and 0.95 for u_1 , the unconditional percentile of X_1 . We can numerically calculate the integral bound of X_1 : $\int_{lb_1}^{ub_1} c(u_1) du_1 = \int_0^{ub_1} c(F(x_1)) du_1 = Pr(X_1 \leq x_1) = F(x_1)$. With the calculated integral bound for X_1 , we can then apply the proposed approach on the following discrete uncertainty X_2 and generate the event structure.

When the preceding uncertainty is discrete, the copula is not differentiable in terms of this discrete uncertainty and we need to apply the discrete version of the conditional percentile (Darsow et al. 1992),

$$Pr(X_2 \leq x_2 | X_1 = x_1) = \frac{\Delta C}{\Delta F_1(x_1)}(F_1(x_1), F_2(x_2)) = \frac{\Delta C}{\Delta u_1}(u_1, u_2),$$

where $\Delta C(u_1, u_2) = C(u_1, u_2) - C(u_1 - \Delta u_1, u_2) = Pr(X_2 \leq x_2 | X_1 = x_1) \Delta u_1$.

The dependent uniform variable $u_2 = Pr(X_2 \leq x_2)$ is the implicit solution of

$$\int_0^{u_1} \int_0^{u_2} c(u_a, u_b) du_a du_b - \int_0^{u_1 - \Delta u_1} \int_0^{u_2} c(u_a, u_b) du_a du_b = Pr(X_2 \leq x_2 | X_1 = x_1) \Delta u_1.$$

which can be calculated numerically. We can then apply the dependent discrete event tree approach by using the inverse marginal transformations to transform the discrete approximation for the standard uniform random u_X into the corresponding discrete approximation for the original random vector $X = F_X^{-1}(u_X)$ where F_X is the marginal distribution function of the variable X .

When the number of uncertainties increases, we iteratively apply the continuous approach and the discrete approach as appropriate for the event structure.

5.2 A Numerical Example with Continuous and Discrete Uncertainties

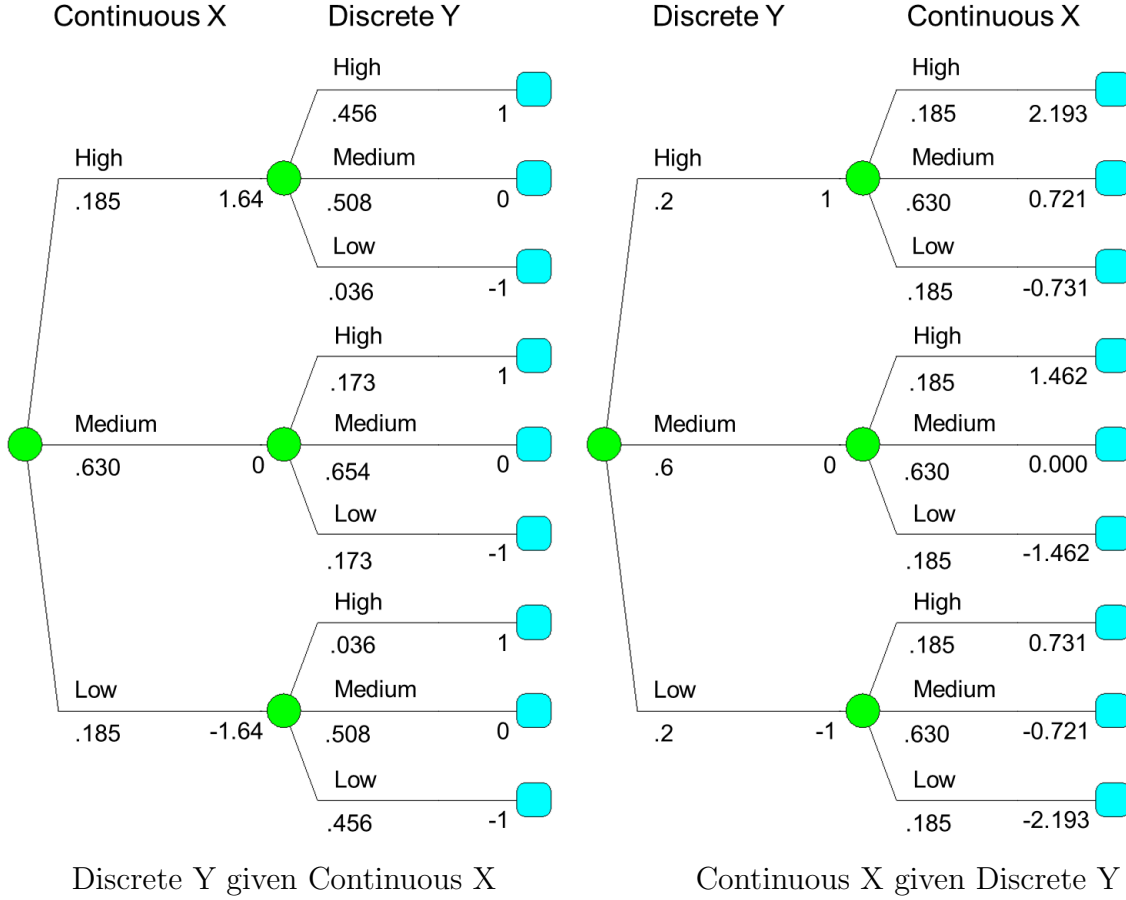
To apply our approach we define two generic uncertainties: a continuous uncertainty X with normal marginal $N(0, 1)$, and a correlated discrete uncertainty Y ($Y = 1$ with probability 0.2; $Y = 0$ with probability 0.6, and $Y = -1$ with probability 0.2). If we assume that the discrete Y is conditioned on the continuous X and use a normal copula with Spearman correlation $\rho = 0.5$, we generate the event tree on the left side of Figure 5. First, we build the discrete approximation for X with the 5th, 50th and 95th percentiles according to the EPT method, which corresponds to the realizations of 0.05, 0.5, and 0.95 for u_1 , the unconditional percentile of X . We then numerically calculate the integral bound of X . For instance, when $X = Low$, we can calculate ub_1 as -0.8965 from $\int_0^{ub_1} c(F(x_1))du_1 = F(x_1) = 0.185$. With the calculated integral bound for X , we can then apply the proposed approach to the subsequent discrete uncertainty Y . For example, the conditional probability $p_{Y=Low|X=Low}$ is calculated as

$$\begin{aligned} p_{Y=Low|X=Low} &= \frac{F(X=Low, Y=Low)}{F(X=Low)} = \frac{\int_{-\infty}^{ub_1} \int_{-\infty}^{\Phi^{-1}(p_2^d)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{p_1} \\ &= \frac{\int_{-\infty}^{-0.8965} \int_{-\infty}^{\Phi^{-1}(0.2)} z_1 z_2 \phi(z_1, z_2, \rho_{12}) dz_1 dz_2}{0.185} = 0.456. \end{aligned}$$

If we exchange the order of the uncertainties X and Y in the event tree and maintain the same normal copula and Spearman $\rho = 0.5$ the process is modified as follows and generates the event tree on the right side of Figure 5. For example, the uniform variable u_X for the conditional scenario $X = Low|Y = Low$ is numerically calculated from $C(u_X, u_Y) - C(u_X, u_Y - \Delta u_Y) = Pr(X \leq Low|Y = Low)\Delta u_Y$ as 0.014, where u_Y and Δu_Y are the percentile of Low and the difference of the two adjacent percentiles (Low and Medium) for discrete uncertainty Y respectively, and $Pr(X \leq Low|Y = Low)$ is the conditional probability 0.185 following the EP-T method. We can then calculate

$X = F_X^{-1}(u_X) = \Phi^{-1}(0.014) = -2.193$, where F_X is the marginal distribution function of the continuous variable X .

Figure 5: Combining Discrete and Continuous Distributions
(Normal copula and Spearman's Rho $\rho = 0.5$)



6 Comparison of Discrete and Continuous Copula-Based Methods

In this section we compare the application of the methodology for discrete uncertainties developed in this paper with the application of the approach in W&D that was designed for

continuous uncertainties. While the two approaches appear similar, there are important differences that result from the technical properties highlighted in Section 5. To motivate the comparison we begin with a discussion of a common practice in decision and risk analyses: discretizing continuous uncertainties before considering the exact distribution and dependence relationships with other uncertainties.

We have described a method for capturing the dependence between discrete uncertainties such as categorical uncertainties (e.g., Success/Failure) and for uncertainties with discrete outcomes (e.g., number of engine failures). Discrete uncertainties have a piecewise CDF so that we can set, for example, $X_i = -1$ if $u_i \leq p_i^d$, $X_i = 0$ if $p_i^d < u_i \leq p_i^d + p_i^m$ and $X_i = 1$ otherwise. However, this approach is not suitable for application to naturally continuous uncertainties that have been discretized. If we apply a discretization method to a continuous uncertainty, each outcome is a representative point of a range of values and the probability associated with the outcome is the probability of a point being selected from its corresponding range.

For example, the EPT method approximates a continuous distribution using a three-point discrete distribution with probabilities 0.185, 0.630, and 0.185 assigned to the 5th, 50th, and 95th percentiles of the continuous distribution. The 5th percentile is the representative value for the lower tail, and the 18.5% probability associated with the 5th percentile is chosen to minimize the error associated with this approximation (Clemen and Reilly, 2013); it is not the probability of an outcome equal to the 5th percentile value. These percentile values and probabilities of the EPT method and other discretization methods are chosen in order to approximate important properties of the corresponding continuous distribution, and typically attempt to match the first, second and perhaps other central moments of the continuous distribution (Hammond and Bickel, 2013).

As a result, the integral bounds are not equal to the associated probability for each discretized outcome. If we apply this method for modeling the dependence among discrete

uncertainties to discretized continuous uncertainties, the conditional distributions will not be moment matching. In contrast, the application of the W&D approach for correlated continuous distributions does result in discrete conditional distributions that are moment matching.

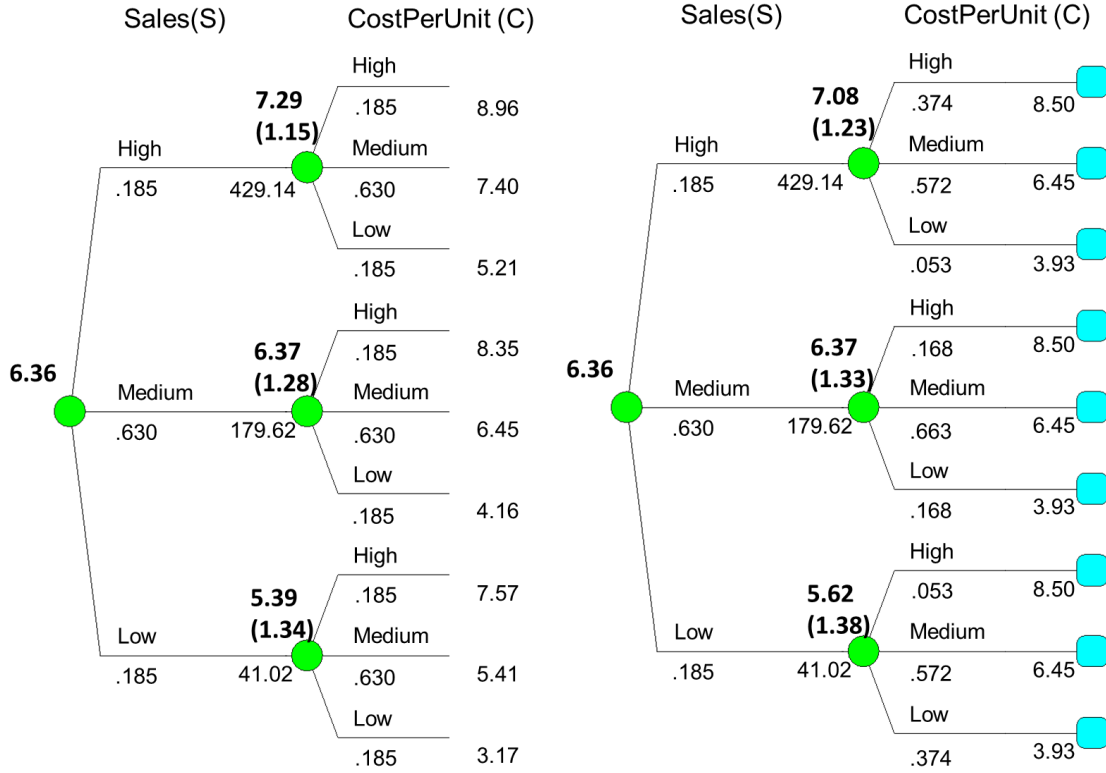
As an example of the implications of applying this copula approach to discretized continuous distributions, consider an analysis that includes the sales of a product (*Sales*) and the costs of the raw materials per unit (*CostPerUnit*) required to manufacture the product. We assume that continuous distributions for *Sales* and *Costs* have been estimated from empirical data analysis or from an expert, and that $Sales \sim Beta(2, 8, 0, 1000)$ and $CostPerUnit \sim Beta(7, 4, 0, 10)$. Then the analyst could apply EPT or any of the standard discretization schemes (see Hammond and Bickel 2013 for a summary) to represent the continuous uncertainty with three (or more) specific outcomes. We will also assume a moderately positive Spearman rank correlation between *Sales* and *CostPerUnit*, $\rho = 0.4$.

The W&D approach uses a copula to capture the correlation between the two continuous uncertainties and then discretizes the resulting continuous marginal and conditional distributions. The left side of Figure 6 shows the EPT discretization of the continuous conditional distributions of *CostPerUnit*; the expected value of *CostPerUnit* is shown above each chance node in bold and the standard deviation is shown in parentheses. There is no closed-form for the conditional distribution of a Beta distribution given the outcome of another Beta, so we used the simulation approach of Clemen and Reilly (1999) and confirmed the W&D moments in the left side of Figure 6 to three decimal places. Thus, the W&D approach will be approximately moment matching, subject to the error in the EPT discretization method used here. .

Notice that the outcomes associated with the conditional discrete distributions of cost per unit change as the 5th, 50th , and 95th percentiles of the corresponding continuous

distributions will be different. However, the probabilities of 0.185, 0.630, and 0.185 stay constant and provide discrete approximations to these conditional continuous distributions that that are approximately moment matching.

Figure 6: Probability of CostPerUnit given Sales



Wang and Dyer (2012) Approach,
Pearson Correlation = 0.4, Normal Copula

Applying Copula-Base Discrete method
on Discretized Uncertainties

An alternative approach would apply the EPT method to Sales and CostPerUnit before considering possible dependence between the two uncertainties. For our assumed distributions, the 5th, 50th and 95th percentiles are 41.02, 179.62, 429.14 and 3.93, 6.45 and 8.50, for Sales and CostPerUnit, respectively. As shown in the right side of Figure 6, this approach results in discrete approximations to the conditional CostPerUnit distributions that reflect their dependence on Sales by changing the probabilities of these same three CostPerUnit levels (3.93, 6.45 and 8.50). The conditional moments do not match for the

two approaches.

For example, $E[\text{CostPerUnit} | \text{Sales} = \text{High}]$ is 7.29 when we apply W&D but it is 7.08 if we discretize the continuous variables before applying the copula; a similar discrepancy holds for the standard deviation of CostPerUnit. These differences will be larger as we increase the absolute value of the correlation between the two uncertainties. This demonstrates that the conditional distributions approximated by discretizing the continuous distributions first and using the copula approach for discrete distributions are not approximately moment matching.

The explanation for the inconsistency lies in the observation that once the dependent distribution has been discretized, there no longer exists a continuous distribution whose moments can be matched. Therefore, the use of the W&D approach should be preferred given correlated continuous uncertainties if a copula is applied to simplify the modeling of dependence.

However, there is a deeper issue. As mentioned previously, common practice in applications of decision analysis is to discretize a continuous variable without first identifying the continuous distribution family and its parameters. For example, we could choose three levels of the Sales uncertainty that represent high, medium and low sales and ask the decision maker to provide the probabilities of those levels of sales. For ease of exposition, we assume that the same levels identified in the right side of Figure 6 are chosen as representative levels for CostPerUnit.

The decision maker is required to estimate the probabilities in Figure 6 for three conditional distributions, one for each assumed level of Sales, but using the same support from the probability mass function. For example, when estimating the top three probabilities in Figure 6, the decision maker should imagine the conditional continuous distribution of CostPerUnit given that Sales are high and then determine the appropriate probabilities for CostPerUnit values of 8.50, 6.45 and 3.93. This requires the decision maker to sub-

jectively estimate the appropriate probabilities for these outcomes that would result in a discrete distribution that is moment matching to the implied continuous distribution, which would be cognitively challenging as discussed by Bansal and Gutierrez (2014). This observation implies that more accurate discrete approximations of conditional distributions are likely to result from the use of continuous distributions and the methodology of W&D rather than by discretizing continuous distributions prior to using both correlations and copulas as we have described, or using subjective assessments of conditional probabilities.

7 Conclusion

Event trees expand rapidly in terms of the number of uncertainties and the number of endpoints to be calculated grows even more quickly, particularly when the event tree is part of a decision tree. This increases the computational burden, but it is also difficult and time consuming to obtain the required conditional probabilities for dependent discrete uncertainties. However, with the correlated event tree methods used here, we only need to assess the marginal distributions and a lower order measure of dependence such as correlation, and we can then calculate the conditional probabilities in closed form. While the ME approach requires the same, smaller number of assessments, copulas can capture a variety of dependent structures as a function of different dependence measures; an example would be tail dependence using a Frank Copula with Kendall's tau.

The W&D approach is only applicable to dependent continuous and differentiable uncertainties and cannot be applied to other situations. The proposed approach was specifically developed to construct a dependent event tree when uncertainties are discrete and can be extended to accommodate combinations of dependent naturally discrete and

continuous uncertainties in the same event tree as illustrated in Section 5. Like the W&D method for continuous correlated uncertainties, this approach is a computationally efficient and practical method to model the under-specified dependence among discrete uncertainties with the use of an event tree. The analyst can reassess one or more of the initial distributions without the cumbersome task of updating all of the conditional probabilities; the copula will make the adjustments.

Another important insight from our work is a novel demonstration of the issues arising from discretizing continuous distributions before accounting for dependence among them. As we show in Section 6, discretizing prior to applying the copula does not lead to moment matching for the conditional distributions. When working with discrete uncertainties it is impossible to provide the required inputs for an EPT discretization; what is the 95th percentile of a coin flip or an engine failure. This increases the difficulty of the assessment task from premature discretization – the need to specify probabilities for fixed supports for the uncertainties – and provides further justification for maintaining continuous uncertainties continuous until dependence has been addressed. While we expect practitioners to continue to discretize uncertainties prior to assessing conditional probabilities or correlations, our analysis provides a methodology to test the implications of this simplification.

Acknowledgments

The authors thank the Associate Editor and anonymous referees for their helpful suggestions and insights. This material is based upon work supported by the Department of Energy and the Lawrence Livermore National Laboratory (Research Grant No. B600310).

References

- Abbas, A. E. 2006. Entropy methods for joint distributions in decision analysis. *IEEE Trans. Engrg. Management* 53(1) 146-159.
- Bansal, S., G. Gutierrez. 2014. Using Imperfectly Elicited Fractiles for the Estimation of Probability Distribution Parameters. Working paper.
- Bethge, M., P. Berens. 2008. Near-maximum entropy models for binary neural representations of natural images. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, 20:97-104. Cambridge, MA: MIT Press.
- Bickel, J. E., J.E. Smith. 2006. Optimal sequential exploration: A binary learning model. *Decision Analysis* 3(1) 16-32.
- Bickel, J. E., J.E. Smith, J.L. Meyer. 2008. Optimal sequential exploration: A binary learning model. *SPE Reservoir Evaluation & Engineering* 11(2) 352-361.
- Bier, V. M., Jr, L.A. Cox. 2007. 15 Probabilistic Risk Analysis for Engineered Systems. *Advances in Decision Analysis: from Foundations to Applications*, 279.
- Budnitz R. J., Apostolakis G., D.M. Boore, L.S. Cluff, K.G. Coppersmith, C.A. Cornell, P.A. Morris. 1998. Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis. *Risk Analysis*. 18 (4): 463-469.
- Cario, M. C., B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Working paper.
- Cherubini, U., E. Luciano, W. Vecchiato. 2004. *Copula Methods in Finance*. John Wiley & Sons, New York.
- Clemen, R. T., G. W. Fisher, R. L. Winkler. 2000. Assessing dependence: Some experimental results. *Management Science*. 46 1100-1115.
- Clemen, R. T., T. Reilly. 1999. Correlations and copulas for decision and risk analysis. *Management Science*. 45 (2), 208-224.
- Clemen, R. T., and Reilly, T. 2013. *Making Hard Decisions*, Cengage Learning.
- Cox, L.A. 2007. Health Risk Analysis for Risk Management Decision-Making. In *Advances in Decision Analysis: From Foundations to Applications*, pp. 325-350. Cambridge University Press : New York.
- Darsow, W. F., B. Nguyen, E. T. Olsen. 1992. Copulas and Markov processes. *Illinois J. Math*. 36 600-642.
- Delfiner, P. 2003. Modeling dependencies between geologic risks in multiple targets. *SPE Reservoir Evaluation & Engineering*, 6(01), 57-64.
- Embrechts, P., F. Lindskog, A. McNeil. 1999. *Modeling Dependence with Copulas and Applications to Risk Management*. Department of Mathematics, ETHZ, Zürich.
- EPRI (Electric Power Research Institute). 2004. Probabilistic Consequence Analysis of Security Threats - A Prototype Vulnerability Assessment Process for Nuclear Power Plants. EPRI Technical Report.
- Ezell, B.C., A.J. Collins. 2010. Response to Parnell, Smith, and Moxley, Intelligent Adversary Risk Analysis: A Bioterrorism Risk Management Model, Risk Analysis, *Risk Analysis*. 30 (6): 875-875.
- Hammond, R.K., J.E. Bickel. 2013. Reexamining Discrete Approximations to Continuous Distributions. *Decision Analysis* 10 (1,) 6-25.

- IAEA. 2010. Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants, IAEA Safety Standards Series No. SSG-3.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. *Physical review*. 106(4), 620.
- Jouini, M. N., R. T. Clemen. 1996. Copula models for aggregating expert opinions. *Operations Research*. 44(3) 444-457.
- Keefer, D. L., S. E. Bodily. 1983. Three-point approximations for continuous random variables. *Management Science*. 29(5) 595-609.
- Keefer D.L. 2004. Underlying event model for approximating probabilistic dependence among binary events. *IEEE Trans. Engrg. Management*. 51(2) 173-182.
- Kullback S, R.A. Leibler. 1951. On information sufficiency. *Ann. Math. Statist*. 22(1):79-86.
- Larralde, H. 2012. Maximum-entropy distributions of correlated variables with prespecified marginals. *Physical Review E*. 86 (6), 061117.
- Levenson, N. 1995. SafeWare: System safety and computers. Addison-Wesley, Reading, MA. p704.
- Lunn, A.D., S.J. Davies. 1998. A Note on Generating Correlated Binary Variables. *Biometrika*. 85(2) 487-490.
- Macke, J. H., P. Berens, A.S. Ecker, A.S. Tolias, M. Bethge. 2009. Generating spike trains with specified correlation coefficients. *Neural Computation*. 21(2), 397-423.
- Merrick, J. R. W., van Dorp, J.R., T. Mazzuchi, J. Harrald, J. Spahn, M. Grabowski. 2002. The Prince William Sound risk assessment. *Interfaces*. 32, 25-40.
- Montiel, L.V., J. E. Bickel. 2012. A Simulation-Based Approach to Decision Making with Partial Information. *Decision Analysis*. 9(4) 329-347.
- Montiel, L.V., J. E. Bickel. 2013. Approximating Joint Probability Distributions Given Partial Information. *Decision Analysis*. 10(1) 26-41.
- Oman, S.D., D. M. Zucker. 2001. Modelling and generating correlated binary variables. *Biometrika*. 88(1) 287-290.
- Papathomas, M. 2008. Correlated Binary Variables and Multi-level Probability Assessments. *Scandinavian Journal of Statistics*. 35 169-185.
- Parnell, G.S, C.M Smith, F. Moxley. 2010. Intelligent Adversary Risk Analysis: A Bioterrorism Risk Management Model. *Risk Analysis*. 30, no. 1 32-48.
- Paté-Cornell, M. E., P. S. Fischbeck, 1994. Risk management for the tiles of the space shuttle. *Interfaces*. 24(1), 64-86.
- Sarin R.K. 1979. An approach for long term forecasting with an application to solar electric energy. *Management Science*. 25(6) 543-554.
- Sklar, A. 1959. *Fonctions de répartition a n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris 8 : 229-231.
- Stabell C.B. 2000. Alternative approaches to modeling risks in prospects with dependent layers. SPE 63204, SPE Annual Tech. Conf. Exhibition, Dallas.
- Wang, T. and Dyer, J. S. 2012. A Copulas-Based Approach to Modeling Dependence in Decision Trees. *Operations Research*. 60 225-242
- Wurbs, R. A. 2005. Comparative evaluation of generalized river/reservoir system models. College Station: Texas Water Resources Institute
- Yi, W., V. M. Bier, 1998. An application of copulas to accident precursor analysis. *Management Science*. 44, 257-270.