Statistical Dependence through Common Risk Factors: With Applications in Uncertainty Analysis

J. René van Dorp[1]

The George Washington University

Summary & Conclusions – **A model for building statistical dependence between marginal distribution with bounded support is discussed. The model is geared towards elicitation of dependence parameters through expert judgment. The resulting joint distribution may be useful in uncertainty analyses where dependence between random variables with a bounded support is present due to common risk factors, such as e.g. in the classical Project Evaluation and Review Technique (PERT).**

Keywords – **Uncertainty Modeling, Mixture of Uniform Distributions, Expert Judgment.**

1. INTRODUCTION

"*The concepts of dependence permeates the Earth and its inhabitants in a most profound manner. Examples of interdependent meteorological phenomena in nature and interdependence in the medical social, and political aspects of our existence, not to mention the economic structures are too numerous to be cited individually" – Drouet and Kotz (2001).* The quote above expresses the need for modeling of dependence between uncertain phenomena. Dependent uncertainty analysis is usually performed with a generic software platform (@Risk, Crystall Ball) or with specialized programs such as UNICORN (see Cooke (1995), Bedford and Cooke (2002) and Kurowicka and Cooke (2002)) or the Probability Bounds Analysis Software by Ferson (1997).

[1] J. René van Dorp, Department of Engineering Management and Systems Egineering, School of Engineering and Applied Science, The George Washington University, 1776 G Street, N.W., Suite 110, Washington D.C. 20052. E-mail: dorpjr@gwu.edu.

The long-standing issue of dependence between random variables has recently been discussed in application areas such as project risk analysis (see, e.g. Duffey and Van Dorp (1998)), accident probability analysis (see, e.g., Yi and Bier (1998)), Finance (see, e.g., Härdle et al. (2002)) and decision analysis (see, e.g., Clemen and Reilly (1999)). Frees and Valdez (1998) introduced dependence in actuarial modeling. These authors unanimously suggested the copula approach (see, e.g. Sklar (1959), Genest and McKay (1986) and Nelsen (1999)) for dependence modeling. An advantage of the copula approach is that it utilizes the decomposition principle by separately describing the uncertainty aspect via the marginal distributions and dependence features between components via copula's.

Although, by now high dimensional sampling routines between a large number of random variables, say a 100 or more, is computationally not too difficult, the representation or modeling of dependence in models of that size in a meaningful manner is still quite cumbersome. With n specified random variables with known marginal distributions, building dependence usually requires specification of $\binom{n}{2}$ correlations (see, e.g., Law and Kelton (1991)). Applications with 100 random variables or more are feasible (see, e.g., Palisade Corporation (1997)), but specification of some $\binom{100}{2} = 4950$ correlations or more becomes a formidable task. Making this task even more daunting is that data bases typically collect information at the individual random variable level, thereby not allowing for the assessment of correlations by means of classical statistical techniques. Hence, one is often compelled in models of this size to utilize the relaxed assumption of independence between the random variables or resort to a probability bounds analysis as suggested by Ferson (2001).

Instead, one may develop an approach to model statistical dependence between the random variables by identifying common risk factors as the source of dependence. The idea of *common risk factors* or *common causes* is not new and has already found wide appreciation in fault tree analysis for chemical and nuclear power plants (see, e.g., Haasl et al. (1981) or Zhang (1989)). Alternatively, common risk factors may be viewed as latent variables. Latent variable models have found wide application in the behavioral sciences (see, e.g., Bartholomew (1987)). Duffey and van Dorp (1998) proposed eliciting dependence via expert judgment by using such common risk factors, however, only a single risk factor was allowed to influence the uncertainty distribution of a random variable which seems too restrictive for practical purposes. The dependence model herein extends the work in Duffey and van Dorp (1998) by allowing multiple common risk factors to affect a single random variable. The extension utilizes a mixture of uniform random variables and its cumulative distribution function to allow for the above mentioned copula approach. A significant reduction is achieved in the required number of dependence parameters compared to the correlation matrix approach (600 in a dependence model with 5 common risk factors and 100 random variables) while allowing separate specification of marginal distributions.

In Section 2, a model for building multivariate dependence between random variables utilizing common risk factors will be discussed. The multivariate dependence of Section 2 utilizes a bivariate dependence model which is discussed in Section 3. In addition, Section 3 introduces a new dependence measure that in its interpretation resembles the well known R^2 measure in regression analysis. The models discussed in Section 2 and 3 allow for elicitation of dependence parameters through the use of expert judgment in a meaningful manner. Section 4 discusses a theoretical result related to the dependence model in Section 2. In Section 5, the model is applied to a PERT example (see, e.g., Winston (1993)). In the example, the effect of neglecting dependence will be benchmarked against a longstanding controversy regarding the use of beta distributions and triangular distributions in PERT analyses (see e.g. Clark (1962), Grubbs (1962) and Kamburowski (1997)) with an assumption of independence between the random variables. Table 1 below summarizes the analysis results. A small project network consisting of 18 activities (see Figure 1) and its accompanying minimal completion time was used to compare the effect of a mild dependence assumption amongst the durations of these activities against an existing controversy regarding the type of distribution that should be used to model duration uncertainty (combined with an independence assumption). The distributions that were used in the PERT analysis were triangular distributions (suggested by Johnson (1997)), a four parameter beta distributions (following a method suggested by Malcolm et al. (1959)) and the Two-Sided Power (TSP) distribution, a recent extension of the triangular distribution and suggested by Van Dorp and Kotz (2002).

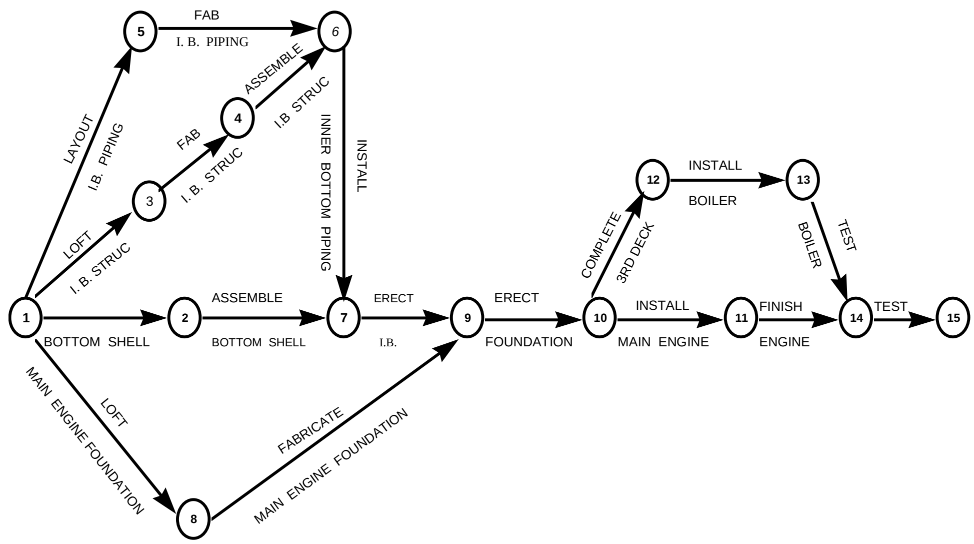

Figure 1. Example Project Network $\mathcal{P}$ for Production Process.

	Mean	St. Dev.
Triangular - Independence	155.15	5.04
Beta - Independence	150.01	4.06
TSP $(n = 5)$ - Independence	149.85	2.96
Triangular - Mild Dependence	154.92	8.74

Table 1. Mean and Standard Deviation of the Project Completion Time Distribution using Triangular, Beta and TSP $(n = 5)$ under Independence and Triangular distributions under Dependence.

Note that the standard deviation of the project completion time practically doubles in case of a mild dependence assumption (fourth row in Table 1) when compared to standard deviations regarding the existing controversy of using a triangular or beta distribution (first and second row in Table 1). Finally, Section 6 provides some concluding remarks.

2. MULTIVARIATE DEPENDENCE MODEL

Figure 2 displays the influence diagram representing the multivariate dependence model between random variables $X_j, j = 1, \ldots, n$. An aggregate risk factor in Figure 2 is a combined measure of risk for a particular random variable arising from multiple common risk factors.

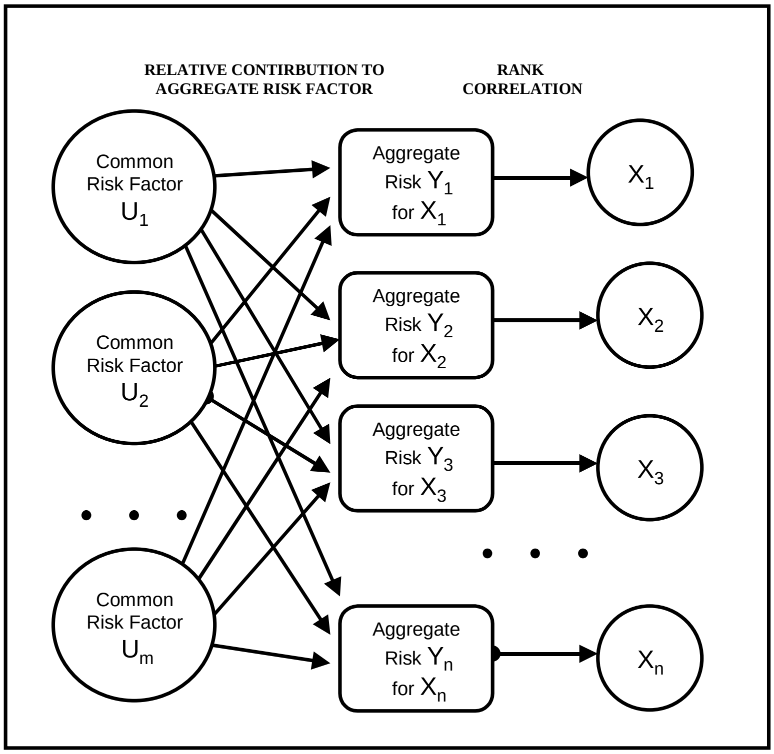

Figure 2. A Model for Statistical Dependence between Random Variables due to Common Risk Factors

Similar to latent variables, common risk factors may not have a natural attribute scale, such as, e.g. Engineering Change Orders (ECO's) and different common risk factors may be measured on different scales. In light of these constraints it is suggested to follow the uniform latent variable approach (see, e.g., Bartholomew (1987)), namely to model common risk factors as independent uniform latent random variables $U_i, i = 1, \ldots, m$, where the lowest risk level for risk factor i is transformed to 0 and the highest to 1. Transforming different risk factors to uniform latent variables allows for elicitation of tradeoff weights $w_i, i = 1, \ldots, m$ for each risk factor through expert judgment utilizing, for example, psychological scaling methods (See, e.g., Cooke (1991)) or the Analytical Hierarchy Process (AHP) method (Saaty, (1980). Figure 3 displays an example pairwise comparison question in the context of the example discussed in Section 5 when utilizing the AHP process.

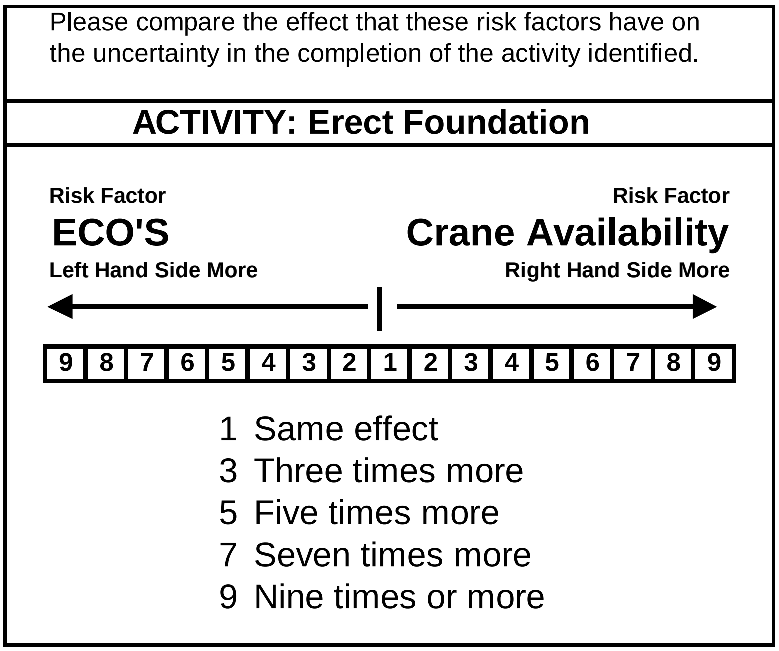

Figure 3. Example AHP Question in the Context of the Example in Section 5 for Eliciting Trade Off Weights between Risk Factors.

The proposed aggregate risk factor Y for random variable X is then calculated as a weighted linear combination

$$Y = \sum_{i=1}^{m} w_i U_i, \sum_{i=1}^{m} w_i = 1,\ w_i \geq 0, \tag{1}$$

where $U_i, i = 1, \ldots, m$ are as mentioned above. For each random variable X_j in Figure 1, relative contributions of the common risk factors $w_{ij},\ i = 1,\ \ldots, m$ need to be specified to aggregate risk from the common risk factors. In addition, a dependence parameter between X_j and the random variable's aggregated risk factor Y needs to be specified. Hence, the total number of dependence parameters that need to be elicited, equals $m \cdot n + n$. With $m = 5$ and $n = 100$ random variables this amounts to 600 dependence parameters compared to the $\binom{100}{2} = 4950$ correlations in a correlation matrix approach to build dependence between 100 specified marginal distributions. Also, no modifications to the dependence parameters are needed due to possible inconsistencies when expert judgment is used to assign these parameters, as is the case with the correlation matrix approach (see, e.g., Iman and Conover (1982)) utilized by popular software programs such as @Risk (See, Palisade Corporation (1997)).

3. BIVARIATE DEPENDENCE MODEL

The multivariate dependence model in Figure 2 utilizes expression (1) and a bivariate dependence model between a random variable X and its aggregated risk Y. A one parameter copula approach (see, e.g., Genest and Mackay (1986)) will be used for this bivariate dependence model. Although a variety of copulas may be used in Figure 2, it is suggested to use the diagonal band (DB) copula (shown in Figure 4A) with dependence parameter θ, first introduced by Cooke and Waij (1986).

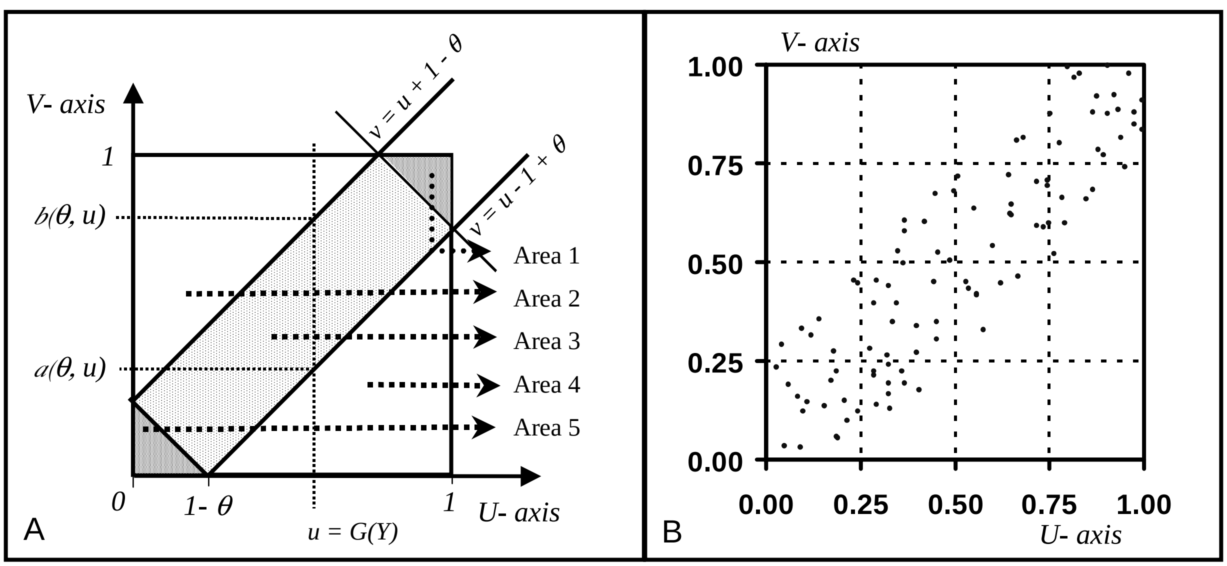

Figure 4. A: DB Copula seen from above

B: Random Sample of size 100 from a DB Copula with $\theta = 0.75$.

The advantage of using the $DB(\theta)$ copula is that utilizing its structure, its single parameter θ may be indirectly elicited via expert judgment and a new dependence measure (to be defined in the next section). In addition, the effect of sampling from the $DB(\theta)$ copula is small compared to, for example, sampling from the Maximal Entropy copula (see, e.g., Van Dorp (1991) and Meeuwissen (1993)) with identical correlation. This is especially true when the latter differences are compared to those associated with an assumption of independence (See, Van Dorp (1991)).

The probability density $d(u,v)$ at the end-points of the diagonal band (Areas 1 and 5 in Figure 4) equals $(1-\theta)^{-1}$ and is exactly twice that of the middle part (Area 3 in Figure 4) $\frac{1}{2}(1-\theta)^{-1}$ and attributes no probability mass above and beneath the diagonal band (Areas 2 and 4 in Figure 4). For illustration, Figure 4B displays a random sample of size 100 generated from a DB-copula with $\theta = 0.75$. From the structure of the DB copula it follows that for $\theta = 0$ $(\theta = 1)$ U and V are independent (identical), while for $0 < \theta < 1$ an intermediate degree of positive dependence is specified (negative dependence may be attained by distributing mass along the second diagonal from $(1,0)$ to $(0,1)$). It is straightforward to calculate the correlation in a DB copula to be

$$\rho(X,Y) = \theta(1 + \theta - \theta^2). \tag{2}$$

To elicit the dependence parameter θ of the DB copula a stepwise approach could be: i) use a direct elicitation approach for the rank correlation and ii) solve for θ. In general, however, correlations are difficult to interpret perhaps suggesting the need for alternative indirect elicitation procedures to determine θ. Clemen and Reilly (1999) and Kraan (2002) discuss a variety of methods for eliciting dependence that use the somewhat intricate statistical concepts like correlation, probability of concordance, joint probability and conditional probability. Kraan (2002) states that for all of these techniques the experts require some training regarding these concepts. When eliciting information from experts, it is desirable to design a meaningful elicitation procedure for engineers so that such information can easily be related to observables (see, e.g., Chaloner and Duncan (1983)). Although the probability concepts above fall within this category, a new dependence measure will be constructed below, exploiting the structure of a DB Copula, that is observable, can be expressed in terms of θ and utilizes only basic statistical concepts such as a simple average and range of a random variable. Note that Spearman's rank correlation $\rho(X,Y)$ given by (2) is not such an *observable* measure.

3.1. Eliciting the DB Copula Parameter

Restricting ourselves to absolute continuous random variables X with bounded support $[a, b]$ and c.d.f. $F(\cdot)$, the concept *range of the support* of X, denoted by $R(X)$, is introduced as

$$R(X) = b - a. \tag{3}$$

The measure $R(X)$ describes the total range where realizations of X can be observed. Next, consider the range of the conditional distribution of $(X|Y = y, \theta)$, i.e. the distribution of X where one knows i) the state of the different common risk factors resulting in aggregate risk Y (cf. (1)) and ii) the dependence parameter θ of the DB copula. Alternatively, we may use $(X|G(Y) = u, \theta)$, where $G(\cdot)$ is the c.d.f. of Y. In the case presented in Figure 4A, it follows that

$$R\{(X|Y = y, \theta)\} = R\{(X|G(Y) = u, \theta)\} = F^{-1}\{b(\theta,u)\} - F^{-1}\{a(\theta, u)\}. \tag{4}$$

From (4) it follows that the average range of the support of the conditional distribution $(X|Y = y, \theta)$, denoted by $R(X|Y, \theta)$, equals

$$R(X|Y, \theta) = \int_{u=0}^{u=1} R\{(X|G(Y) = u,\, \theta)\}du. \tag{5}$$

Utilizing (3) and (5) a meaningful dependence measure

$$\xi(X|Y, \theta) = \left(1 - \frac{R(X|Y, \theta)}{R(X)}\right)100\%. \tag{6}$$

is introduced. The dependence measure $\xi(X|Y, \theta)$ and its interpretations resemble the popular R^2-measure in regression analysis. The measure $\xi(X|Y, \theta)$ may be interpreted as *the average percent* $reduction$ *in the range* of X by knowing the state of the common risk factors. Similarly to Pearson's product moment correlation it can be shown that $\xi(X|Y,\theta)$ is invariant under linear transformations of X (see, e.g., Joag-Dev (1984)). However, unlike Pearson's product moment correlation, the behavior of $\xi(X|Y,\theta)$ is dependent on the form of the marginal distribution c.d.f. $F(\cdot)$ of X and the dependence parameter θ, but not dependent on the structural form of the c.d.f. $G(\cdot)$ of the risk factor Y. In contrast, it is well known that the Spearman's rank correlation $\rho(X,Y)$ (cf. (2) for the $DB(\theta)$ distribution) between X and Y is invariant under all non-decreasing transformations of X and Y and is thus not dependent on either the marginal form of $F(\cdot)$ nor that of $G(\cdot)$ (see, e.g., Joag-Dev (1984)). Hence we conclude, at least conceptually, that the dependence measure $\xi(X|Y,\theta)$ given by (6) lies somewhere in between Pearson's product moment correlation and Spearman's rank correlation.

Figure 5 studies the effect of the marginal form of the c.d.f. F on the resulting dependence parameter θ and the corresponding rank correlation $\rho(X,Y)$ (cf. (2)) as a function of a specified value for $\xi(X|Y,\theta)$. Figure 5A, displays the required level of θ and corresponding rank correlation (on the y-axis) to achieve a particular average % explanation $\xi(X|Y,\theta)$ for both a symmetric $Triang(0,0.5,1)$ and a $Beta(0,1,2\frac{1}{2},2\frac{1}{2})$ distribution (cf. (25) with $n=2$ and cf. (23), respectively). It can be shown that the mean and the variance of a $Triang(0,0.5,1)$ and $Beta(0,1,2\frac{1}{2},2\frac{1}{2})$ are identical. Observe from Figure 5A, that the corresponding values for θ and the rank correlation are practically identical. The same holds for Figure 5B, where the analysis involves a skewed $Triang(0,0,1)$ distribution and a skewed $Beta(0,1,1,2)$ distribution, again with identical means and variances. From Figure 5A and 5B it may be observed that the dependence parameter θ and corresponding rank correlation seems to primarily be affected by the mean and variance of the c.d.f. $F(\cdot)$. (Recall that the forms of the density of a $Triang(0,0.5,1)$ and $Beta(0,1,2\frac{1}{2},2\frac{1}{2})$ are quite different.) Figure 5C compares a similar analysis for the symmetric $Triang(0,0.5,1)$ and skewed $Triang(0,0,1)$ distributions, the latter one having a larger variance measure than the former. Finally, Figure 5D presents analogous results for the symmetric $Triang(0,0.5,1)$ and $Uniform(0,1)$ distributions, the latter one also having a larger variance measure than the former. From Figure 5C and 5D it may be observed, at least empirically, that the larger the uncertainty in the marginal distribution $F(\,\cdot\,)$ of X, the larger the dependence has to be in the $DB(\theta)$ copula to achieve the same average % explanation $\xi(X|Y,\theta)$ in $F(\,\cdot\,)$ by the common risk factor Y.

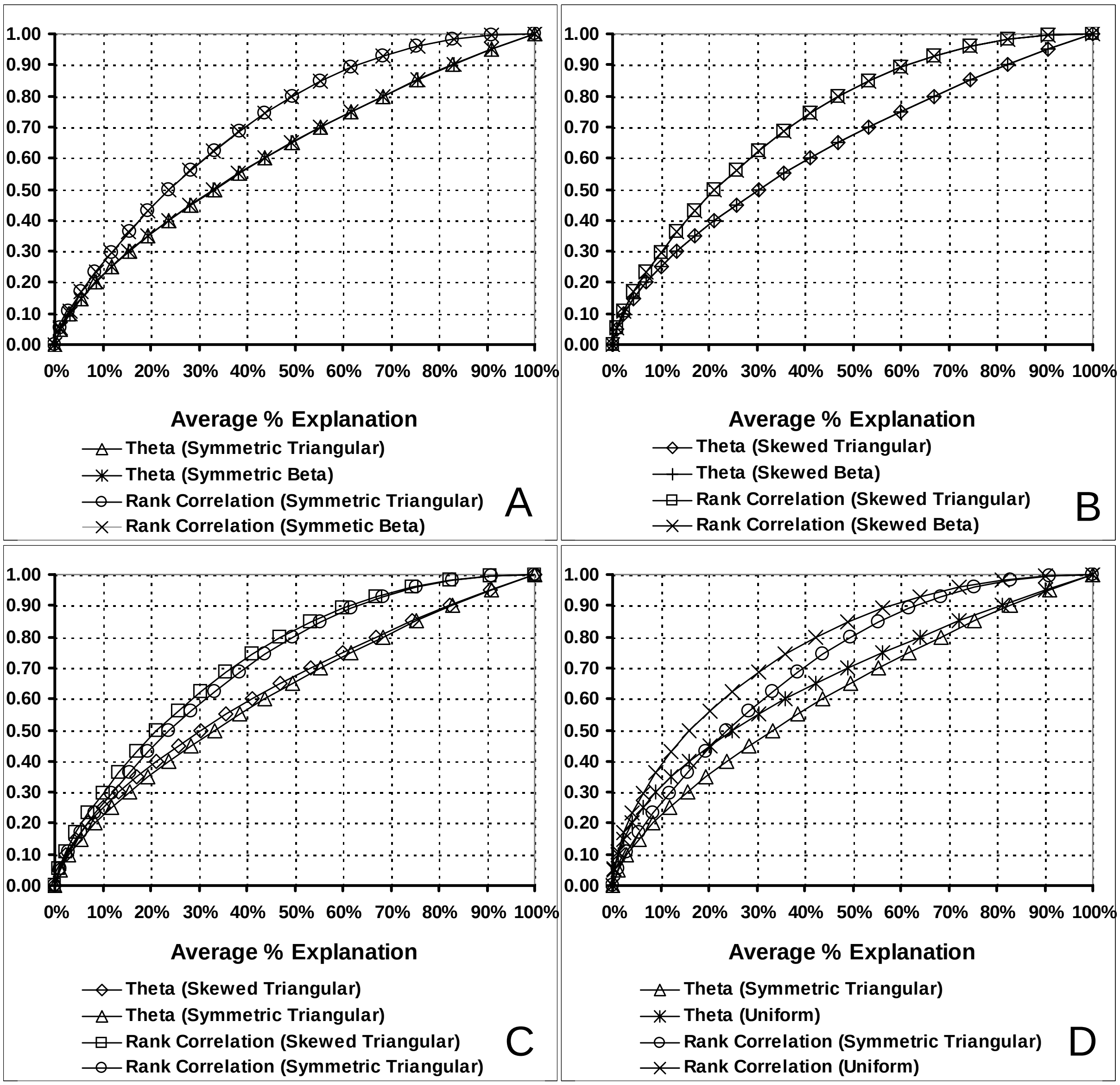

Figure 5. Relationship between Average % Explanation, Diagonal Band parameter θ and rank correlation between risk factor and marginal distribution

A: $Triang(0,\frac{1}{2},1)$ and $Beta(0,1,2\frac{1}{2},2\frac{1}{2})$; B: $Triang(0,0,1)$ and $Beta(0,1,1,2)$;

C: $Triang(0,\frac{1}{2},1)$ and $Triang(0,0,1)$; D: $Triang(0,\frac{1}{2},1)$ and $Uniform(0,1)$;

Suppose, for example, that for a particular activity duration with risk factor Y being "*The Weather*", a range of 10 days has been assessed without knowing the state of the weather. The average percent reduction $\xi(X|Y,\theta)$ may next be elicited from an expert by asking a question of the following type:

"*Not knowing the state of the weather a spread of 10 days has been assessed for the activity. Suppose you knew the state of the weather during the completion of the activity, on average within a spread of how many days could you now assess the completion of this activity?*".

If the expert answers 5 days, $\xi(X|Y,\theta)$ (cf. (6)) corresponds to 50%. In other words, 50% of the original uncertainty of the activity durations is explained by knowing the weather. Note that the question is formulated in terms of the range of the original random variable X, which is $observable$. In case of the marginal distribution $X \sim Triang(0,5,10)$ or $Beta(0,10,2\frac{1}{2},2\frac{1}{2})$ we conclude from Figure 5A, utilizing the invariance of $\xi(X|Y,\theta)$ under linear transformations, that $\theta \approx 0.65$ (and $\rho \approx 0.80$) in the $DB(\theta)$ copula to achieve such an average % explanation $\xi(X|Y,\theta)$ by Y. On the other hand, if $X \sim Triang(0,0,10)$ or $Beta(0,10,1,2)$ it follows from Figure 5B that $\theta \approx 0.67$ (and $\rho \approx 0.82$). Finally, if $X \sim Uniform[0,1]$ it follows from Figure 5D that $\theta \approx 0.71$ (and $\rho \approx 0.85$).

Regardless of the function form of $F(\cdot)$ it follows from the structure of the $DB(\theta)$ copula that

$$\begin{cases} \xi(X|Y,\theta) = 0\% & \theta = 0 \\ \xi(X|Y,\theta) = 100\% & \theta = 1, \end{cases} \tag{7}$$

and values of $\theta \in (0,1)$ result in values of $\xi(X|Y,\theta) \in (0,1)$. If $F(\cdot)$ is known in closed form, as in the case of the triangular distribution, $\xi(X|Y,\theta)$ may be expressed in terms of θ in a closed form. However, if $F(\cdot)$ is not known in closed from, as in the case of a beta distribution, it follows from (7) that a bisection method (see, e.g., Press et al. (1989)) may be designed to solve for θ up to a desirable level of accuracy (higher than, for example, the accuracy achieved by utilizing Figure 5 and interpolation techniques). Such a bisection method is described in the appendix in Pseudo Pascal (and may also be used when $F(\,\cdot\,)$ is available in a closed form).

As suggested by one of the referees, the range of support method above (which is connected to the diagonal band copula) has intuitive appeal to experts working with triangular distributions and four parameter beta distributions, which are used in a PERT context, but not elsewhere, and may thus be considered a limitation. However, applications of the beta distribution go far beyond that of the PERT context including the fields of, e.g., Ecology, Reliability and Statistical Quality Control, spawning the publication of a separate handbook entitled "Beta Distributions and its Applications" and edited by Gupta and Nadarajah (2003). The use of the triangular distribution is particularly popular in Monte Carlo Software programs such as @Risk (See Palisade Corporation (2000)), Arena (See, Kelton et al. 2002) and Crystal Ball (See Decision Engineering, Inc. (2003)) with additional application areas such as e.g. business and finance, engineering design and environmental assessment. National Energy Board (1998) is an example of a publication in the environmental assessment arena utilizing triangular distributions. In addition, recently, the choices of families of distribution with bounded support have been enhanced by the discovery of the two-sided power distributions and their generalizations (see, Van Dorp and Kotz (2002, 2003a)) and generalized trapezoidal distributions (see, Van Dorp and Kotz (2003b)). In my opinion, most applied phenomena and their uncertainty are of a bounded nature.

However, if one were to prefer the use of univariate distributions with unbounded support and different copulas than the diagonal band copula, only a minor modification of the range of support method above is needed. Instead of range of support as defined by (3) the concept $(1-\alpha)$% credibility range of a random variable X would need to be introduced given by

$$R_\alpha(X) = F^{-1}(1-\frac{\alpha}{2}) - F^{-1}(\frac{\alpha}{2}), \tag{8}$$

using for example popular values for α such as $0.01,\ 0.05$ or 0.10. In addition, $(4),\ (5)$ and (6) would need to be modified accordingly, yielding

$$R_\alpha\{X|Y=y,\theta\} = R_\alpha\{X|G(Y)=u,\,\theta)\} = F^{-1}_{(X|Y=y,\theta)}(1-\frac{\alpha}{2}) - F^{-1}_{(X|Y=y,\,\theta)}(\frac{\alpha}{2}), \tag{9}$$

$$R_\alpha(X|Y,\theta) = \int_{u=0}^{u=1} R_\alpha\{X|G(Y)=u,\ \theta)\}du, \tag{10}$$

and

$$\xi_\alpha(X|Y,\theta) = \left(1 - \frac{R_\alpha(X|Y,\theta)}{R_\alpha(X)}\right)100\%. \tag{11}$$

In (9), $F_{(X|Y=y,\,\theta)}(x)$ is the conditional c.d.f. of the random variable X, were one to know the state of aggregate risk Y and the parameter θ of the diagonal band distribution. Expressions (8), (9), (10) and (11) allow for straightforward generalization to other copulas than the diagonal band copula in Figure 4.

The measure $\xi_\alpha(X|Y,\theta)$ (cf. (11)) may be interpreted as *the average percent $reduction$ in the* $(1-\alpha)$*% credibility range* of a random variable X by knowing the state of the common risk factors. However, the inclusion of the parameter α in the dependence measure $\xi_\alpha(X|Y,\theta)$ (cf. (11)) requires an additional level of cognitive processing when using this measure to elicit dependence via expert judgment and therefore looses intuitive appeal. Combining the latter with the observation that the effect of sampling from the $DB(\theta)$ copula is small compared to, for example, sampling from the Maximal Entropy copula (see, e.g., Van Dorp (1991) and Meeuwissen (1993)) with identical correlation, leads me to prefer the more intuitive measure $\xi(X|Y,\theta)$ given by (6). The Maximal Entropy copula is the most natural copula given two marginal distributions and a correlation constraint (see, Meeuwissen (1993)).

4. DISTRIBUTION OF A LINEAR COMBINATION OF UNIFORM VARIABLES

To use the copula approach to model bivariate dependence between a random variable X and its aggregated risk Y, both X and Y need to be transformed to the uniform marginals U and V of the copula. The required (integral) transformations of X and Y are $F(X)$ and $G(Y)$, where the functions $F(\cdot)$ and $G(\cdot)$ are the c.d.f.'s of X and Y, respectively. For a known marginal distribution for random variable X, $F(\cdot)$ is readily obtained either in closed form (e.g., in the case of a triangular distribution) or through numerical routines (e.g., for a beta distribution). The c.d.f. of the linear combination Y (cf. (1)) is given by

$$G(y) = Pr(Y \leq y) = \sum_{v_1=0}^{1} \ldots \sum_{v_m=0}^{1} (-1)^{\sum_{i=1}^{m} v_i} \left\{ \frac{(y - \sum_{i=1}^{m} w_i v_i)^m}{m! \prod_{i=1}^{m} w_i} \right\} 1_{[0,\infty)}(y - \sum_{i=1}^{n} w_i v_i). \quad (12)$$

(see, Mitra (1971) or Barrow and Smith (1979)). Unfortunately their proofs $-$ geared towards mathematically oriented readers $-$ are very concise and somewhat difficult to follow. The proof discussed in the next section which seems to be new, is geometric in nature and is based on the time honored inclusion-exclusion principle

$$Pr\Big\{ \bigcup_{i=1}^{m} A_i \Big\} = \sum_{i=1}^{m} Pr(A_i) - \sum\sum_{i<j} Pr\{A_i \cap A_j\} + \quad (13)$$
$$\sum\sum\sum_{i<j<k} Pr\{A_i \cap A_j \cap A_k\} - \ldots + (-1)^m Pr\Big\{ \bigcap_{i=1}^{m} A_i \Big\},$$

for arbitrary events $A_1, \ldots, A_n$ (not necessarily disjoint) (see, e.g., Feller (1990)). The geometric nature of the proof allows for an efficient algorithm for evaluation of (12) needed for its application in Monte Carlo based uncertainty analyses. The Appendix describes the algorithm in Pseudo Pascal .

4.1. Theoretical Result

Let $C^m = \{\underline{u} \mid 0 \leq u_i \leq 1\}$ be the unit hyper cube in $\mathbb{R}^m$. Let $\underline{v} = (v_1, \ldots, v_m),\ v_i \in \{0, 1\}$ be a vertex (or corner point) of the unit hyper cube C^m and define the simplex $S_{\underline{v}}(y)$ at the vertex $\underline{v}$ as

$$S_{\underline{v}}(y) = \{ \underline{u} \mid \sum_{i=1}^{m} w_i u_i \leq y, u_i \geq v_i, i = 1, \ldots, m \}, \quad (14)$$

where $w_i \geq 0$, $\sum_{i=1}^{m} w_i = 1$. For example, Figure 6A displays C^3 and the simplex $S_{(0,0,0)}(y_1)$ (cf. (14)).

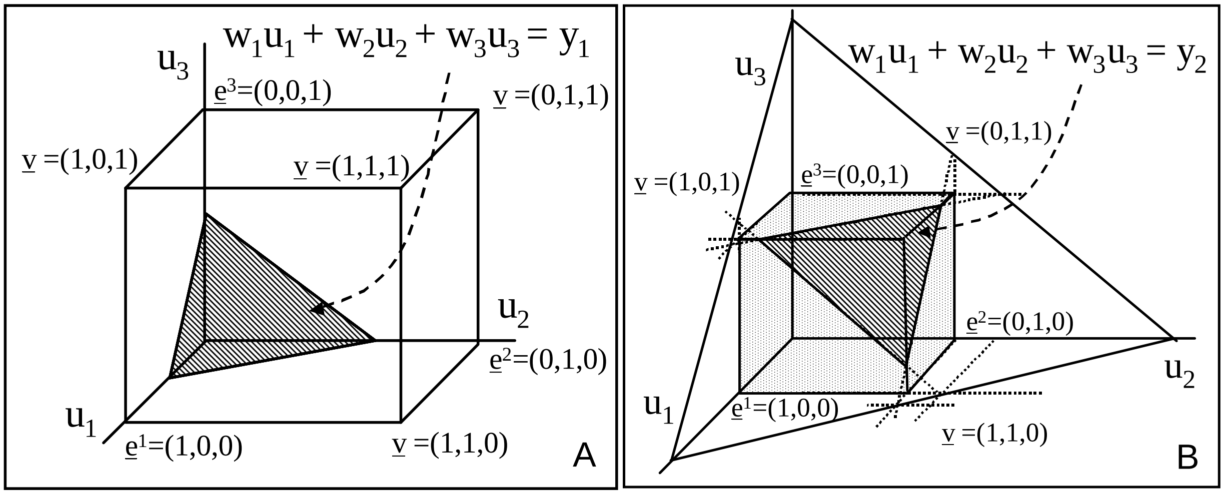

Figure 6. A: Evaluating $G(y_1) = Pr(Y \leq y_1)$ (cf. (12)) for $m = 3$;

B: Evaluating $G(y_2) = Pr(Y \leq y_2)$ (cf. (12)) for $m = 3$.

Note that, for this particular value $0 < y_1 < 1$ only the simplex $S_{\underline{0}}(y)$ at the origin $\underline{0} = (0, 0, 0)$ is a non-empty set since $\underline{0} = (0, 0, 0)$ is an element of the half space $\{ \underline{u} \mid \sum_{i=1}^{m} w_i u_i < y_1\}$. When the value of y increases in (14), additional corner points $\underline{v}$ of the unit hyper cube will join the half space $\{ \underline{u} \mid \sum_{i=1}^{m} w_i u_i < y_1\}$ resulting in additional non-empty simplices at those points. For example, consider Figure 6B for a particular value $0 < y_2 < 1,\ y_2 > y_1$. In Figure 6B, we may recognize the simplex $S_{\underline{0}}(y)$ at the origin $\underline{0}$ as the largest one. In addition, we can observe the three smaller simplices $S_{(1,0,0)}(y_2),\ S_{(0,1,0)}(y_2)$ and $S_{(1,0,0)}(y_2)$ at the corner points $\underline{e}^1 = (1, 0, 0)$, $\underline{e}^2 = (0, 1, 0)$ and $\underline{e}^3 = (0, 0, 1)$, respectively, of approximately equal size. Finally, the three smallest simplices in Figure 6B (indicated with dotted lines) are $S_{(1,1,0)}(y_2)$, $S_{(1,0,1)}(y_2)$ and $S_{(0,1,1)}(y_2)$ at the corner points $(1, 1, 0),\ (1, 0, 1)$ and $(0, 1, 1)$, respectively. No simplex can be observed at the eighth corner-point $(1, 1, 1)$ in Figure 6B since $(1, 1, 1)$ is an element of the half space $\{ \underline{u} \mid \sum_{i=1}^{m} w_i u_i > y_2\}$. Our proof of (12) utilizes the hypervolume of the simplices defined by (14).

***Lemma* 1 :** *The hyper volume* $V\{S_{\underline{v}}(y)\}$ *of the simplex* $S_{\underline{v}}(y)$ *given by* (14) *equals*

$$\frac{(y-\sum_{i=1}^{m} w_i v_i)^m}{m!\prod_{i=1}^{m} w_i} \cdot 1_{[0,\infty)}(y-\sum_{i=1}^{n} w_i v_i). \tag{15}$$

***Proof* :** From the definition of (14) it immediately follows that for $y \geq 0$

$$V\{S_{\underline{0}}(y)\} = \int_{u_1=0}^{1}\int_{u_2=0}^{1-\sum_{i=1}^{1}\frac{w_i}{y}u_i} \ldots \int_{u_m=0}^{1-\sum_{i=1}^{m-1}\frac{w_i}{y}u_i} du_m \ldots du_1. \tag{16}$$

Changing the variables of integration to $z_i = w_i u_i / y, i = 1, \ldots, m$, the integral in (16) becomes

$$V\{S_{\underline{0}}(y)\} = \frac{y^m}{\prod_{i=1}^{m} w_i}\int_{z_1=0}^{1}\int_{z_2=0}^{1-\sum_{i=1}^{1} z_i} \ldots \int_{z_m=0}^{1-\sum_{i=1}^{m-1} z_i} dz_m \ldots dz_1. \tag{17}$$

The integral in (17) is the hyper-volume of the unit simplex

$$S = \{\, \underline{u} \mid \sum_{i=1}^{m} u_i \leq 1, u_i \geq 0, i = 1, \ldots, m\,\}. \tag{18}$$

Realizing that the Dirichlet distribution (see, e.g., Kotz et al. (2000)) with density function

$$\frac{\Gamma(\eta)}{\prod_{i=1}^{m+1}\Gamma(\eta \cdot \nu_i)} \cdot \left(\prod_{i=1}^{m}(u_i)^{\eta \cdot \nu_i - 1}\right)\left(1-\sum_{i=1}^{m} u_i\right)^{\eta \cdot \nu_{m+1}} \tag{19}$$

where $\eta > 0,\ \nu_i > 0,\ \sum_{i=1}^{m+1}\nu_i = 1$ has support S (cf. (18)) and by setting its parameters in (19) equal to $\eta = m+1,\ \nu_i = \frac{1}{m+1},\ i = 1, \ldots, m+1$ it immediately follows from (17), (19) and the fact that $S_{\underline{0}}(y) = \emptyset$ for $y < 0$ that

$$V\{S_{\underline{0}}(y)\} = \frac{y^m}{m!\prod_{i=1}^{m} w_i} \cdot 1_{[0,\infty)}(y). \tag{20}$$

Again changing variables $x_i = u_i - v_i,\ i = 1, \ldots, m$ we arrive utilizing (14) at

$$V\{S_{\underline{v}}(y)\} = V\{S_{\underline{0}}(y - \sum_{i=1}^{n} w_i v_i)\}. \tag{21}$$

The lemma now follows from (20) and (21). $\square$

***Theorem* 1 :** *The c.d.f. of the weighted linear combination* Y *given by* (1), *where* U_i *are independent* $[0,1]$ *uniform random variables is given by* (12).

Proof : The support of Y follows from (1) as $[0,1]$. Let $\underline{0} = (0,\ldots,0)$ be the origin vertex of the unit hyper cube C^m and let $\underline{e}^i = (e_1,\ldots,e_m),\ i = 1,\ldots,m,$ be the unit vertices of C^m (See, Figure 6), i.e. $e_i = 1, e_j = 0,\ j = 1,\ldots,m, j \neq i$. For illustration we shall consider the case $m = 3$ and evaluating $Pr(Y \leq y_1)$ for the value of $0 < y_1 < 1$ indicated by Figure 6A and that of $Pr(Y \leq y_2)$ for the value of $y_2 > y_1,\ 0 < y_2 < 1,$ depicted in Figure 6B. Figure 6A displays C^3 and $S_{\underline{0}}(y_1)$ (cf. (14)). Figure 6B displays C^3, $S_{\underline{0}}(y_2), S_{\underline{e}^1}(y_2),\ S_{\underline{e}^2}(y_2)$ and $S_{\underline{e}^3}(y_2)$. From (1) and the independence of $U_i,\ i = 1,\ldots,m$ it follows that in Figure 6A
$Pr(Y \leq y_1) = V\{S_{\underline{0}}(y_1)\}$. In Figure 6B the calculation of $Pr(Y \leq y_2)$ is somewhat more complicated. Figure 6B shows that

$$Pr(Y \leq y_2) = V\{S_{\underline{0}}(y_2)\} - V\Big\{\bigcup_{i=1}^{3} S_{\underline{e}^i}(y_2)\Big\}. \tag{22}$$

Note that (22) also holds for the value y_1 in Figure 6A as $S_{\underline{e}^i}(y_1) = \emptyset,\ i = 1,\ldots 3$. Generalizing to $\mathbb{R}^m$ we obtain directly

$$Pr(Y \leq y) = V\{S_{\underline{0}}(y)\} - V\Big\{\bigcup_{i=1}^{m} S_{\underline{e}^i}(y)\Big\} \tag{23}$$

The inclusion-exclusion principle (cf. (13)) yields

$$V\Big\{\bigcup_{i=1}^{m} S_{\underline{e}^i}(y)\Big\} = \sum_{i=1}^{m} V\{S_{\underline{e}^i}(y)\} - \sum\sum_{i<j} V\{S_{\underline{e}^i}(y) \cap S_{\underline{e}^j}(y)\} + \sum\sum\sum_{i<j<k} V\{S_{\underline{e}^i}(y) \cap S_{\underline{e}^j}(y) \cap S_{\underline{e}^k}(y)\} - \ldots + (-1)^m V\Big\{\bigcap_{i=1}^{m} S_{\underline{e}^i}(y)\Big\}. \tag{24}$$

Utilizing (14) it follows that the intersections of the simplices $S_{\underline{e}^i}(y)$ in (24) are all of following form

$$\bigcap_{i \in I} S_{\underline{e}^i}(y) = S_{\underline{v}}(y) \tag{25}$$

where $I \subset \{1, \ldots, m\}$ and $\underline{v} = \sum_{i \in \mathrm{I}} \underline{e}^i$. For example, $S_{(1,1,0)}(y_2)$ in Figure 6B is the intersection of $S_{(1,0,0)}(y_2)$ and $S_{(0,1,0)}(y_2)$. From (25), (24) and (23) we conclude that

$$Pr(Y \leq y) = \sum_{v_1=0}^{1} \ldots \sum_{v_m=0}^{1} (-1)^{\sum_{i=1}^{m} v_i} V\{S_{\underline{v}}(y)\}. \tag{26}$$

The proof of the theorem follows from Lemma 1. □

From the proof it follows that an efficient method to evaluate the distribution in (12) for a particular value of y and a given set of weights $\underline{w} = (w_1, \ldots, w_m)$ is to develop a recursive algorithm enumerating all vertices $\underline{v}$ of the hypercube C^m and evaluate the hypervolume of the simplex at each vertex $\underline{v}$ given by (15) when a vertex is visited by the procedure. The next section will discuss an application of the dependence model in the PERT domain.

5. EXAMPLE - A CONTROVERSY IN PERT

Johnson (1997) proposed the triangular distribution to be used as an alternative to the beta distribution. Its parameters have a one-to-one correspondence to an optimistic estimate a, a most likely estimate m and a pessimistic estimate b of an activity duration T in a PERT network. Much earlier, Malcolm et al. (1959) fitted a four-parameter $Beta(a, b, p, q)$ distribution

$$\begin{array}{l} f_T(t|a, b, p, q) \ = \ \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \dfrac{(t-a)^{p-1}(b-t)^{q-1}}{(b-a)^{p+q-1}} \\ a \leq t \leq b, \ p > 0, \ q > 0, \end{array} \tag{27}$$

by estimating a, m and b and using the method of moments to overcome difficulties involved with interpreting the beta parameters by setting

$$\begin{cases} E[T] = \dfrac{a+4m+b}{6} \\ Var[T] = \frac{1}{36}(b-a)^2. \end{cases} \tag{28}$$

Solving for the beta parameters using (28) has been controversial (see e.g. Clark (1962), Grubbs (1962)) and its use is still (see e.g. Kamburowski (1997)) subject to a discussion. Van Dorp and Kotz (2002) suggested the use of a Two-Sided Power $TSP(a,m,b,n)$ distribution, an extension of the triangular distribution, defined by the density

$$f_X(x|a,m,b,n) = \begin{cases} \frac{n}{(b-a)}\left(\frac{x-a}{m-a}\right)^{n-1} & a < x \leq m \\ \frac{n}{(b-a)}\left(\frac{t-x}{b-m}\right)^{n-1} & m \leq x \leq b \ , \end{cases} \tag{29}$$

as a proxy to the beta, specifically in problems of assessment of risk and uncertainty (such as in PERT). For $n = 2$ in (29) the TSP density coincides with the density of a triangular distribution. The expressions for the mean and the variance for (29) result in

$$E[X] = \frac{a+(n-1)m+b}{n+1} \tag{30}$$

and

$$Var(X) = (b-a)^2 \cdot \left\{ \frac{n - 2(n-1)\frac{(m-a)}{(b-a)}\frac{(b-m)}{(b-a)}}{(n+2)(n+1)^2} \right\}. \tag{31}$$

From (30) it follows that for a triangular distribution $(n = 2)$ $E[X]$ may over- or under estimate $E[T]$ in (28) depending on whether m is less or greater than the midpoint $(a+b)/2$. However, for a TSP distribution with $n = 5$, the mean values $E[T]$ in (28) and $E[X]$ in (30) coincide. Perhaps more importantly, it follows from (31) that in case of a triangular distribution $(n = 2)$

$$Var[T] = \frac{1}{36}(b-a)^2 < \frac{3}{72}(b-a)^2 \leq Var[X] \leq \frac{1}{18}(b-a)^2. \tag{32}$$

and for a TSP distribution with $n = 5$ we have

$$\frac{1}{84}(b-a)^2 \leq Var(X) \leq \frac{5}{252}(b-a)^2 < Var[T] = \frac{1}{36}(b-a)^2. \tag{33}$$

Hence, from (32) ((33)) it follows that $Var[X]$ of a triangular distribution (a TSP distribution with $n = 5$) is always larger (less) than $Var[T]$ in (28), regardless of the values of a, m and b, which possibly augments the controversy related to the setup given by (28).

With a project network structure between activities, the random variables representing the uncertainty in activity duration and an assumption of independence between these random variables, the uncertainty in the completion time of the project can be obtained using a combination of the Critical Path Method (CPM) (see, e.g., Winston (1993)) and Monte Carlo methods (see, e.g., Vose (1996)). However, the independence assumption is highly suspect for many large engineering projects involving multiple activities of a similar type and/or different activity types which are influenced by common risk factors (see, e.g., Duffey and Van Dorp (1998)). An example of a common risk factor between activities is inclemental weather for e.g. painting, outfitting of piping and electrical systems or other activities scheduled under the "open sky" conditions in the same time period. In this example it will be shown that the effect of ignoring dependence between these activity durations on the project completion time distribution is of the same magnitude as those observed when modeling uncertainty in activities durations via (27) and (28) (beta), (29) with $n = 2$ (triangular) or (29) with $n = 5$ (TSP), respectively.

5.1. Description

The dependence model in Figure 2 and the elicitation methods to be described herein have been applied in Greenberg (1998). Multiple elicitation sessions with naval architects were used to specify: (a) the parameters a, m, b for the uncertainty distribution of 254 activity durations in a PERT network and (b) the parameters for the dependence model in Figure 2 with 5 common risk factors: weather, manning availability, material availability, crane availability and ECO's. A complete description of the case study is presented in Greenberg (1998). We shall demonstrate the approach by means of a smaller example in the PERT domain. Figure 1 in Section 1 shows an 18-activity project network in the ship building domain from Taggart (1980). The uncertainty in each activity duration could be elicited through expert judgment via a lower bound a, most like estimate m and upper bound b as described in Table 2. Modern-day ship production is a manufacturing domain in which innovative design and build strategies require special attention to risk factors that may impact cost and delivery time. Two major risk areas are the impact of ECO's and crane unavailability. Engineering changes may come from a variety of sources -- such as owner-requested changes, inadequate design specifications, interface problems for vendor-furnished equipment, etc. Cranes are used to lift large prefabricated units and their unavailability due to outages may result in substantial project delays. The relative contributions of ECO and crane unavailability to aggregate risk and the percent reduction in range of the completion time of the activities given the state of these risk factors are specified in Table 2.

ID	Activity Name	a	m	b	w_{ECO}	w_{CRANE}	$S(X\|Y,\theta)$
1	Shell: Loft	22	25	30	1	0	25%
2	Shell: Assemble	35	37	43	1	0	25%
3	I.B.Piping: Layout	19	22	29	0.5	0.5	25%
4	I.B.Piping: Fab.	4	5	10	1	0	25%
5	I.B.Structure: Layout	23	26	31	1	0	25%
6	I.B.Structure: Fab.	16	18	24	1	0	25%
7	I.B.Structure: Assemb.	11	14	20	0.5	0.5	25%
8	I.B.Structure: Install	6	7	12	0.5	0.5	25%
9	Mach Fdn. Loft	25	28	33	0.5	0.5	25%
10	Mach Fdn. Fabricate	33	35	40	0.5	0.5	25%
11	Erect I.B.	27	30	37	0.2	0.8	25%
12	Erect Foundation	6	7	11	0.2	0.8	25%
13	Complete #rd DK	4	5	9	0.2	0.8	25%
14	Boiler:Install	6	7	10	0	1	25%
15	Boiler:Test	9	10	15	1	0	25%
16	Engine: Install	6	7	12	0	1	25%
17	Engine: Finish	17	20	26	1	0	25%
18	FINAL Test	13	15	20	1	0	25%

Table 2. Parameters for modeling the uncertainty in activity durations for the project network in Figure 1, Relative Contribution of ECO's (w_{ECO}) and Crane Unavailability (w_{crane}) to aggregate Risk Y and average reduction in range given the state of the common risk factors.

Note that due to similarity in exposure to ECO's and usage of the crane these parameters may not need to vary by activity, thereby further reducing the assessment of dependence parameters by pre-grouping similar activities in terms of reliance on common risk factors. A 25% reduction in range is assumed across the board. This reduction of 25% may be viewed as a mild form of dependence. (Reductions were observed in Greenberg (1998) in the order of 75%).

5.2. Project Completion Time Distribution Analysis

To show the effect of dependence between the activity durations, the minimal completion time distribution of the project in Figure 1 has been generated via Monte Carlo analysis, utilizing the information in Table 2, the dependence model described above and activity durations with a triangular form $($cf. (29) with $n = 2)$. (Amongst the TSP and beta distribution, the triangular distribution is the only one that is completely specified by a, m and b without additional assumptions.) The latter minimal completion time distribution is then compared in Figure 7 with the project completion time distribution assuming independence between the activity durations with a triangular form $($cf. (29) with $n = 2)$, a beta form (via (28) and employing the method of moments) and finally a TSP form $($cf. (29) with $n = 5)$. In addition, the minimal completion time of 144 days of a standard CPM analysis utilizing only the most likely estimates in Table 2 is depicted by a vertical line. The mean and the standard deviation of the project completion distribution for the four combinations are provided in Table 1 in Section 1. It follows from Table 1 that the results involving the standard deviation of the project completion time associated with the independence assumption are consistent with the earlier observation that the variance in the triangular (TSP) distributions are strictly larger (smaller) than those in its beta counter part (see (32) and (33)). With the independence assumption between beta activity durations, the use of (28) results in a significant reduction in the mean of the project completion time and a substantial reduction in its standard deviation when compared to utilizing triangular distributions whose parameters are directly specified by the three estimates $a,\ m$ and b (See Table 2). Hence, the adoption of (28) may not be consistent with a conservative approach towards estimating project completion time and its uncertainty. Note that when utilizing TSP distributions ((29) with $n = 5$), a similar mean shift occurs in the project completion time and even a larger shift in the standard deviation, providing an even more optimistic scenario.

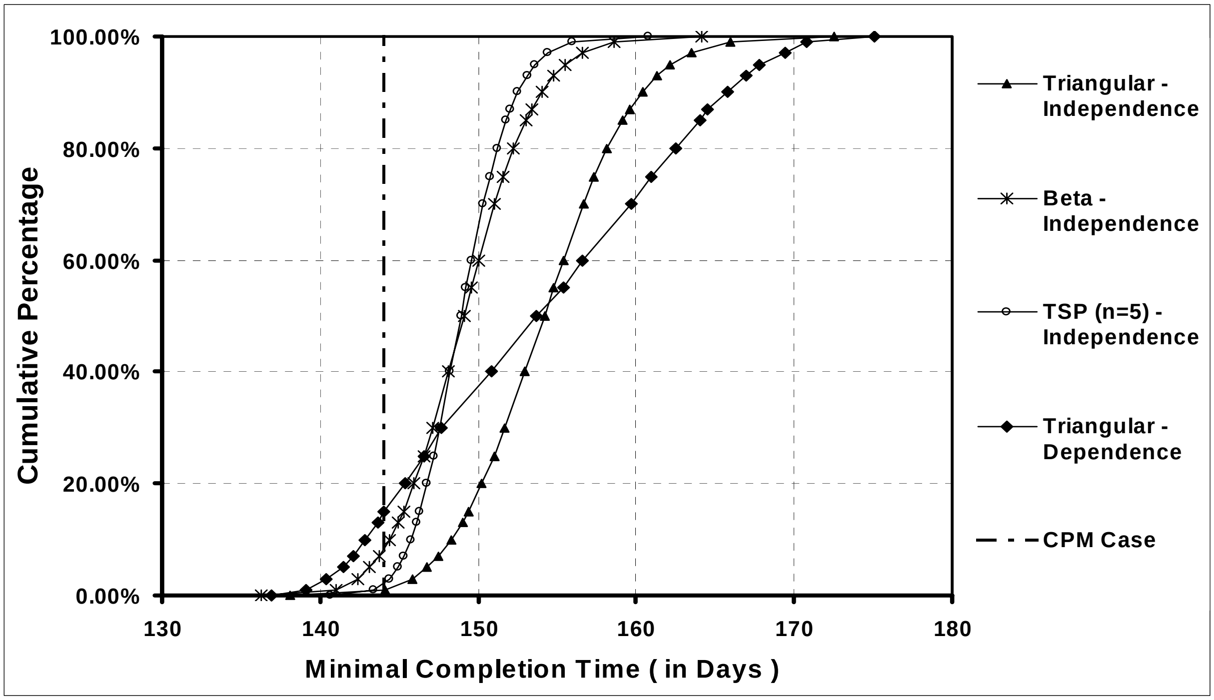

Figure 7. Comparison of Distributions of Minimal Completion Time for the project in Figure 5.

The most notable result in Figure 7 and Table 1, however, follows from comparing the completion time distribution under an assumption of (mild) dependence with the distributions assuming independence. Although no mean shift occurs when comparing the first and fourth rows in Table 1, the standard deviation of the completion time of the project almost doubles. The same observation follows from Figure 7 where the distribution under the dependence assumption possesses a much smaller slope and appears to have a support that overlaps all of its counterparts. Evidently, if the use of (28) and its resulting underestimation of project completion time and uncertainty were a reason for a long standing controversy (see e.g. Clark (1962), Grubbs (1962) and Kamburowski (1997)), it would seem that the issue of modeling dependence deserves similar interest.

Note also that it follows from Figure 7 that the probability of completing the project by 144 days calculated using the standard CPM method is less than 15% regardless of an assumption of dependence or independence. This result is due to the fact that the ingredient distributions of the activity durations are positively skewed. Positively skewed distributions were prevalent in the expert judgment used in Greenberg (1998). Such a prevalence may be explained by the existence of a motivational bias amongst experts resulting in optimism regarding the most likely value of activity completion. This fact could serve as an explanation for a low incidence of project success (on-time) when utilizing standard CPM analysis as a yard stick.

6. CONCLUDING REMARKS

A dependence model has been developed allowing to build a multivariate distribution involving a large number of marginal distributions with bounded support. Dependence parameters of the model may be elicited via expert judgment using indirect elicitation procedures. Via a computational example it has been shown that the effect of an assumption of mild dependence on the minimal completion time of a small project exceeds that of a long standing controversy (over 40 years) regarding the use of triangular or beta distributions in PERT analyses. Although theoretically it is well known that the uncertainty of an output parameter may be greatly under estimated when unjustifiably assuming independence (e.g. think of the well known formula $Var(X+Y) = Var(X) + Var(Y) \; + \; 2Cov(X,Y)$), the benchmarking of the effect of mild dependence against this controversy perhaps argues for even greater attention to modeling dependence in uncertainty analyses, especially when a large number of random variables are involved.

7. ACKNOWLEDGMENTS

I am indebted to Samuel Kotz and Thomas A. Mazzuchi for their valuable comments and encouragement in developing this paper. I would like to express my gratitude to both referees and the Editor for their careful review of the paper and their suggestions which improved the presentation and the content of the first version.

8. REFERENCES

Barrow, D.L., Smith, P.W., 1979. Spline Notation Applied to a Volume Problem, *American Mathematical Monthly*, 86: 50-51.

Bartholomew, D.J., 1987. Latent Variable Models and Factor Analysis, *Oxford University Press*.

Bedford, T.J., Cooke, R.M., 2002. Vines - a new graphical model for dependent variables. *Annals of Statistics*, 30 (4): 1031-1068.

Chaloner, K.M, Duncan, G.T., 1983. Assessment of a Beta Prior Distribution: PM Elicitation, *The Statistician*, 32: 174-181.

Clark, C.E., 1962. The PERT Model for the Distribution of an Activity. *Operations Research*, 10: 405-406.

Clemen, R., Reilly, T., 1999. Correlations and Copulas for Decision and Risk Analysis. *Management Science*, 45: 208-224.

Cooke, R.M., Waij, R., 1986. Monte Carlo Sampling for Generalized Knowledge Dependence, with application to Human Reliability, *Risk Analysis*, 6 (3): 335-343.

Cooke, R.M., 1991. Experts in Uncertainty, *Oxford University Press*, New York.

Cooke, R.M., 1995. UNICORN, Methods and Code for Uncertainty Analysis, *Atomic Energy Authority* for ESRA, Risley (UK).

Decision Engineering, Inc. 2000. Crystal Ball, Denver, Colorado.

Drouet, D., Kotz, S., 2001. Correlation and Dependence**,** Imperial Press, Singapore.

Duffey, M.R., Van Dorp, J.R., 1998. Risk Analysis for Large Engineering Projects: Modeling Cost Uncertainty for Ship Production Activities, *Journal of Engineering Valuation and Cost Analysis*, 2: 285-301.

Feller, W., 1990. Introduction to Probability Theory and Its Application, Vol. 1. John Wiley & Sons, Incorporated, 3rd edition.

Ferson, S., 1997. Probability bound analysis software, Computing in Environmental Resource Management, Proceedings of the Conference, A. Gertler (ed.), Air and Waste Management Association and the U.S. Environmental Protection Agency, Pittsburgh, Pennsylvania, pp. 669-678.

Ferson, S., 2001. Probability bounds analysis solves the problem of incomplete specification in probabilistic risk and safety assessments, Risk-Based Decision Making in Water Resources IX, Y.Y. Haimes, D.A. Moses and E.Z. Stakhiv (eds.), American Society of Civil Engineers, Reston, Virginia, pp. 173-188.

Frees, E.W., Valdez, E.A., 1998. Understanding Relationships using Copula's. *North America Actuarial Journal*, 2: 1-25.

Genest C., Mackay J., 1986. The Joy of Copulas, Bivariate Distributions With Uniform Marginals, *The American Statistician*, 40 (4): 280-283.

Greenberg M., 1998. An Activity Network Simulation Approach for Estimating the Production Schedule of a Convertible Container Ship, *Masters Thesis*, Engineering Management Department, The George Washington University.

Grubbs, F.E., 1962. Attempts to validate certain PERT statistics or a 'Picking on PERT', *Operations Research*, 10: 912-915.

Gupta, A.K., Nadarajah, S. eds. 2003. Beta Distribution and its Applications, Marcel Dekker, New York, To Appear.

Haasl, D.F., Roberts N.H. , Vesely W.E., Goldberg F.F., 1981. Fault Tree Handbook, Systems and Reliability Research, Office of Nuclear Regulatory Commission, Washington D.C.

Härdle W., Kleinow, T. , Stahl, G., 2002. Applied Quantitative Finance, Theory and Computational Tools, Springer, e-book.

Iman, R.L., Conover W.J., 1982. A Distribution-Free Approach to Inducing Rank Order Correlation among Input Variables, *Communications in Statistics-Simulation and Computation*, 11 (3): 311-334.

Joag-Dev, K., 1984. Measures of Dependence, *Handbook of Statistics*, 4, Elsevier, North Holland, pp. 79-88.

Johnson, D.G., 1997. The Triangular Distribution as a Proxy for the Beta Distribution in Risk Analysis. *The Statistician.* 46 (3): 387-398.

Kamburowski, J., 1997. New Validations of PERT Times. *Omega, International Journal of Management Science*, 25 (3): 323-328.

Kelton, D.W., Sadowski, R.P., Sadowski, D.A. 2002. Simulation with Arena, McGraw-Hill, New York.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Continuous Multivariate Distributions, Volume 1: Models and Applications, 2nd Edition, *Wiley*, New York.

Kraan, B., 2002. Probabilistic Inversion in Uncertainty Analysis and related Topics, Dissertation, Delft University of Technology, Delft, Holland. ISBN 90-9015710-7.

Kurowicka, D., Cooke, R.M. 2002. The Vine Copula Method For Representing High Dimensional Dependent Distributions: Application To Continuous Belief Nets, Proceedings of the 2002 Winter Simulation Conference, E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds., pp. 270-278.

Law, A.M., Kelton, W.D., 1991. Simulation Modeling and Analysis, *Mc-Graw Hill*, 2nd ed., New York.

Malcolm, D.G., Roseboom, C.E., Clark, C.E., Fazar, W., 1959. Application of a Technique for Research and Development Program Evaluation. *Operations Research,* 7: 646-649.

Meeuwissen, A.M.H., 1993. Dependent Random Variables in Uncertainty Analysis. Dissertation, Delft University of Technology, Delft, Holland. ISBN 90-9006471-0.

Mitra, S.K., 1971. On the Probability Distribution of The Sum of Uniformly Distributed Random Variables, *SIAM Journal of Applied Mathematics,* 20 (2): 195-198.

National Energy Board 1998. Probabilistic Estimae of Hydrocarbon Volumens in the Mackenzie Delta and Beuafort Sea Discoveries. Calgary, Alberta.

Nelsen, R.B., 1999. An Introduction to Copulas. Springer, New York.

Palisade Corporation, 1997. @RISK, Advanced Risk Analysis for Spreadsheets, *Palisade Corporation,* Newfield, New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vettering, W.T., 1989. *Numerical Recipes in Pascal*, Cambridge University Press, Cambridge.

Saaty, T.L., 1980. The Analytic Hierarchy Process, NY, McGraw Hill.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris,* 8: 229-231.

Taggart, R., 1980. Ship Design and Construction, *The Society of Naval Architects and Marine Engineers* (SNAME), New York.

Van Dorp, J.R., 1991. Dependence Modeling for Uncertainty Analysis. *TWAIO Report*, Delft University of Technology.

Van Dorp, J.R., Kotz, S., 2002. A Novel Extension of the Triangular Distribution and its Parameter Estimation, *The Statistician*, 51 (1): 63-79.

Van Dorp, J.R., Kotz, S., 2003a. Generalizations of Two Sided Power Distributions and their Convolution. *Communications in Statistics: Theory and Methods*, 2 (9): 1703 - 1723.

Van Dorp, J.R., Kotz, S. 2003b. Generalized Trapezoidal Distributions, *Metrika*, 58 (1): 85 - 97.

Vose D., 1996. Quantitative Risk Analysis, A Guide to Monte Carlo Simulation Modeling, *Wiley,* New York.

Winston, W.L., 1993. Operations Research, Applications and Algorithms, *Duxbury Press*.

Yi, W., Bier, V., 1998. An application of Copulas to Accident Precursor Analysis, *Management Science*, 44: S257-S270.

Zhang, Q., 1989. A General Method Dealing with Correlations in Uncertainty Propagation in Fault Trees, *Reliability Engineering and System Safety*, 26: 231-247.

9. APPENDIX

The procedure $CalcTheta(\boldsymbol{\theta}, \xi, N)$ below is a bisection algorithm that calculates the bandwidth parameter θ of the DB copula given a value for the dependence measure $\xi\,(X|Y,\theta)$ (cf. (6)). The procedure $CalcTheta$ uses the procedure $AverageReduction(\boldsymbol{\xi},\ \theta, N)$ to calculate the value for $\xi\,(X|Y,\theta)$ given a value for dependence parameter θ and discretization accuracy N. The procedure $AverageReduction$ uses a function $F^{-1}(\,\cdot\,)$ for the inverse of the c.d.f. of the random variable X.

$AverageReduction(\boldsymbol{\xi},\ \theta, N);$

$Step\,1: Sum:\,=0;\ i:\,=1;$

$Step\,2: u:\,=\frac{i}{N}$

$Step\,3: a:\,=Max(0, u-1+\theta);\ b:\,=Min(u+1-\theta, 1)$

$Step\,4: Sum:\,=Sum+\frac{F^{-1}(b)-F^{-1}(a)}{N-1}$

$Step\,5: If\, i<N-1\, then\ i:\,=i+1;\ Goto\, STEP\,2;$

$Step\,6: \xi:\,=100\cdot\left(1-\frac{Sum}{F^{-1}(1)-F^{-1}(0)}\right)$

$CalcTheta(\boldsymbol{\theta}, \xi, N)$

$Step\,1:\ If\ (\xi=0\%\,)\ or\ (\xi=100\%)\ then\ \theta_k:=\xi;\ Goto\, STEP\,7;$

$Step\,2: d:\,=0; e:\,=1;$

$Step\,3: \theta_k:\,=\frac{d+e}{2}; AverageReduction(\xi_k,\ \theta_k, N)$

$Step\,4: If\,|\xi\text{-}\xi_k|<\delta\, then\, Goto\, STEP\,7$

$Step\,5: If\,\xi_k>\xi\ then\, e:\,=\theta_k\, else\ d:\,=\theta_k;$

$Step\,6: Goto\, STEP\,3$

$Step\,7:\ \theta:\ =\theta_k;$

The procedure $CalcCDF(\boldsymbol{G},y,m,\underline{w})$ below evaluates the c.d.f. of Y given by (1) by making a call to the recursive procedure $VisitVertices(\boldsymbol{G},y,i,\underline{v},m,\underline{w},\ \Pi)$. The algorithm uses functions $ProductWeights(\underline{w},m)$ to calculate $\Pi=\prod_{i=1}^{m}w_i,\ SumElements(\underline{v},m)$ to calculate $\Sigma=\sum_{i=1}^{m}v_i$ and $SumProducts(\underline{v},\underline{w},m)$ to calculate $\psi=\sum_{i=1}^{m}w_iv_i$.

$VisitVertices(\boldsymbol{G},y,i,\underline{v},m,\underline{w},\ \Pi);$

$Step\ 1: if\, i<m\, then$

$v_i:\ =0;\ VisitVertices(\boldsymbol{G},y,i,\underline{v},sum,m,\underline{w},\ \Pi);$

$v_i:\ =1;\ VisitVertices(\boldsymbol{G},y,i,\underline{v},sum,m,\underline{w},\ \Pi);$

$Step\,2:\Sigma:\ =SumElements(\underline{v},m);\psi:\ =SumProducts(\underline{v},\underline{w},m);$

$Step\,3: If\ (y-\psi)>0\,then\ G:\ =G+(-1)^{\Sigma}\frac{y-\psi}{m!\cdot\Pi}$

$CalcCDF(\boldsymbol{G},y,m,\underline{w});$

$Step\,1: If\ y\leq 0\ then\,G:\ =0;Stop;$

$Step\,2: If\ y\geq 1\,then\,G:\ =1;\ Stop;$

$Step\,3:\Pi:\ =ProductWeights(\underline{w},m);$

$Step\,4: VisitVertices(\boldsymbol{G},y,1,\underline{v},m,\underline{w},\ \Pi);$