A Distribution for Modeling Dependence Caused by Common Risk Factors

J. René van Dorp
The George Washington University, Washington D.C., USA

Submitted to: European Safety and Reliability 2003 Conference Proceedings
Submitted : December 2002   Revised: March 2003

ABSTRACT: The cumulative distribution function (cdf) of a finite mixture of independent uniform random variables will be derived. The distribution is useful for uncertainty analyses in application domains such as, e.g., project risk analysis, decision analysis, finance, accident probability analysis and actuarial analysis, particularly when dependence between uncertain elements is present due to common risk factors. Use of the distribution reduces the number of dependence parameters that need to be assessed to specify dependence when compared to a correlation matrix approach. An example discussing the effect of dependence in the project risk analysis domain utilizing the mixture distribution is presented.

1 INTRODUCTION

The motivation for the construction of the cumulative distribution function (cdf) of a finite mixture of uniform random variables arose in the development of a risk analysis approach for project networks (see, Van Dorp & Duffey (1999)). The duration of the activities in such a network may be modeled as random variables. With the project network structure and an assumption of independence between these random variables the completion time of the project can be readily obtained using a combination of the Critical Path Method (see, e.g. Winston (1993)) and Monte Carlo methods (see, e.g. Vose (1996)). However, the independence assumption between these random variables may be specious (e.g. Duffey & Van Dorp (1998)) and one may resort to modeling statistical dependence between the random variables utilizing e.g. a correlation matrix approach via multivariate normal distributions. 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)), decision analysis (see, e.g., Clemen and Reilly (1999)) and actuarial modeling (see, e.g., Frees and Valdez (1998)).

With n pre-specified marginal distributions a correlation matrix approach generally requires the specification of $\binom{n}{2}$ correlations. In project risk analysis project sizes of 100 activities are not uncommon and specification of $\binom{100}{2} = 4950$ correlations becomes a formidable task. When using engineering judgment to specify such a correlation matrix, inconsistencies occur as the correlation matrix needs to be positive definite, and often modifications to the engineering judgment are needed (e.g. Iman & Conover (1982)). Instead, one may develop an approach to model statistical dependence between these activity durations using common risk factors. 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)). Examples of possible common risk factors in a project risk analysis context are; weather, engineering change orders, productivity of workforce etc.

The dependence model in Van Dorp & Duffey (1999) uses common risk factors for modeling dependence, but is restricted to 1 common risk factor per disjoint subsets of activities. This, however, implies that only 1 risk factor influences the uncertainty in an activity duration which may be too restrictive for practical purposes. The cdf to be derived in this paper allows multiple common risk factors to influence the uncertainty in an activity duration and is thus more flexible than the dependence model in Van Dorp & Duffey (1999).

Figure 1 displays the influence diagram representing the relaxed dependence model.

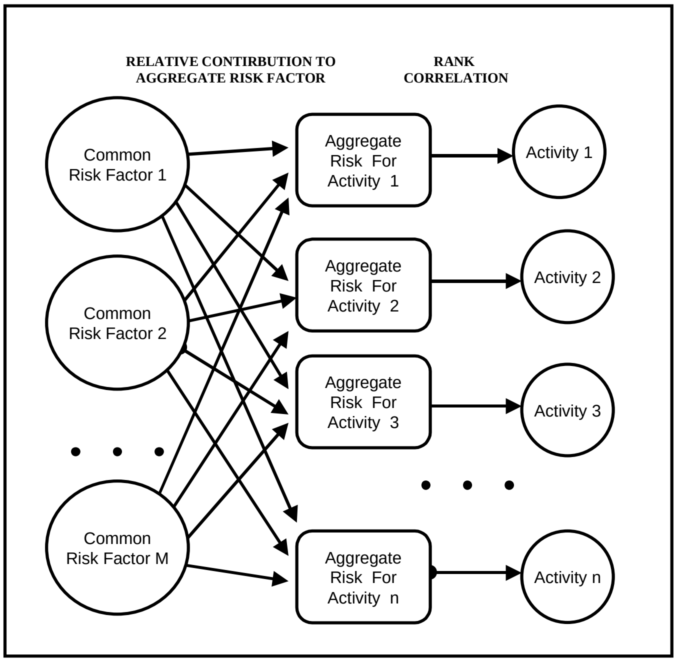

Figure 1. A Model for Statistical Dependence between Activity Durations due to Common Risk Factors

The authors Frees and Valdez (1998), Duffey and Van Dorp (1998), Yi and Bier (1998), Clemen and Reilly (1999) and Härdle et al. (2002)) 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. A one parameter copula is used to model the dependence between an activity duration and its associated aggregate risk factor indicated in Figure 1. The correlation between the uniform marginals of the copula is the rank correlation between the aggregate risk factor and its activity duration. The rank correlation has been proposed as an appropriate measure of positive statistical dependence (see, e.g. Joag-Dev (1984)).

An aggregate risk factor in Figure 1 is a combined measure of risk for a particular activity duration arising from the common risk factors between activities. A common risk factor may not have a natural attribute scale and different common risk factors are generally measured on different scales. Therefore, for convenience, each common risk factor $i, i=1,\ldots,m$ is modeled as a uniform latent random variable U_i, where the lowest risk level for such a risk factor is transformed to 0 and the highest risk is transformed to 1. Latent random variable models have found wide application in the behavioral sciences (see, e.g. Bartholomew (1987)). Transforming the different risk factors to the same scale allows engineers to tradeoff risk factors using a tradeoff elicitation approach like e.g. *swing weights* as described in Clemen (1995). Using such an elicitation approach, relative contributions w_i of each risk factor pertaining to a particular activity may be obtained. The proposed aggregate risk factor Y for an activity is then calculated as

$$Y=\sum_{i=1}^{m} w_i U_i. \tag{1}$$

Note that Y is a mixture of uniform random variables U_i with mixture weights w_i, $i=1,\ldots,m$. To be able to use the copula approach to model statistical dependence between an activities aggregated risk and the activity duration as in Figure 1, the random variable Y needs to be transformed into a uniform random variable. It is well known that the required transformation is $F(Y)$, where F is the cdf of the mixture of uniform random variables given by (1). The cdf of Y will by derived in the next section.

Note that for each activity m relative contributions of individual risk factors need to be specified to aggregate risk, in addition to a rank correlation between an activity's aggregate risk factor and the activity duration. Hence, the total number of parameters that needs to be specified, equals $m \cdot n + n$. With 5 common risk factors and 100 activities in a project network this amounts to 600 parameters as opposed to the 4950 correlations in a correlation matrix approach to build dependence between pre-specified marginal distributions. Also, no modifications to the dependence parameters are needed due to inconsistencies when engineering judgment is used to specify these parameters. The dependence model in Figure 1 has been successfully tested in Greenberg (1998). Multiple elicitation sessions with Naval Architects were used in Greenberg (1998) to specify; (1) the parameters for the uncertainty distribution of 254 activity durations in a project network and (2) the parameters for a dependence model as in Figure 1 with 5 common risk factors.

2 A MIXTURE 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 cdf'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 cdf of the linear combination Y (cf. (1)) is given by

$$Pr(Y \leq y) = G(y) = \sum_{v_1=0}^{1} \dots \sum_{v_m=0}^{1} \tag{2}$$

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

(see, Mitra (1971) or Barrow & 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

$$Pr\Big\{ \bigcup_{i=1}^{m} A_i \Big\} = \sum_{i=1}^{m} Pr(A_i) - \tag{3}$$
$$\sum\sum_{i<j} Pr\{A_i \cap A_j\} +$$
$$\sum\sum\sum_{i<j<k} Pr\{A_i \cap A_j \cap A_k\} - \dots$$
$$+ (-1)^m Pr\Big\{ \bigcap_{i=1}^{m} A_i \Big\},$$

for arbitrary events $A_1, \dots, A_n$ (not necessarily disjoint). The geometric nature of the proof allows for an efficient algorithm to evaluate (2) and needed for the application of (2) in Monte Carlo based uncertainty analyses. The Appendix describes the algorithm in Psuedo Pascal .

2.1 Theoretical Result

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

$$S_{\underline{v}}(y) = \tag{4}$$
$$\{ \underline{u} \mid \sum_{i=1}^{m} w_i u_i \leq y, u_i \geq v_i, i = 1, \dots, m \},$$

where $w_i \geq 0$, $\sum_{i=1}^{m} w_i = 1$. Figure 2A displays C^3 and the simplex $S_{(0,0,0)}(y_1)$ (cf. (4)). Figure 2B displays C^3 and $S_{(0,0,0)}(y_2)$, $S_{(1,0,0)}(y_2)$, $S_{(0,1,0)}(y_2)$, $S_{(1,0,0)}(y_2)$, $S_{(1,1,0)}(y_2)$, $S_{(1,0,1)}(y_2)$ and $S_{(0,1,1)}(y_2)$. Our proof of (2) utilizes the following lemma.

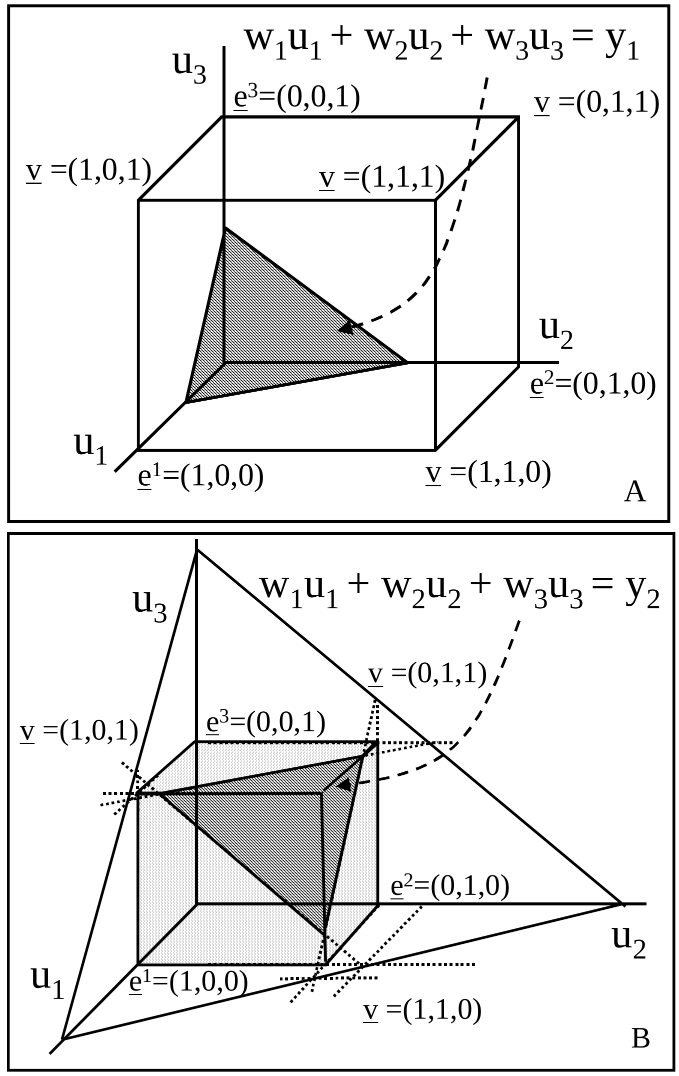

Figure 2. A: Evaluating $G(y_1)$ for $m = 3$; B: Evaluating $G(y_2)$ for $m = 3$.

***Lemma* 1 :** *The hyper volume* $V\{S_{\underline{v}}(y)\}$ *of the simplex* $S_{\underline{v}}(y)$ *given by* (4) *equals*

$$\frac{(y - \sum_{i=1}^{m} w_i v_i)^m}{m! \prod_{i=1}^{m} w_i} \cdot 1_{[0,\infty)}(y - \sum_{i=1}^{n} w_i v_i). \tag{5}$$

***Proof* :** From the definition of (4) it immediately follows that for $y \geq 0$

$$V\{S_{\underline{0}}(y)\} = \tag{6}$$
$$\int_{u_1=0}^{1} \int_{u_2=0}^{1-\sum_{i=1}^{1} \frac{w_i}{y} u_i} \dots \int_{u_m=0}^{1-\sum_{i=1}^{m-1} \frac{w_i}{y} u_i} du_m \dots du_1.$$

Changing the variables of integration to $z_i = w_i u_i / y$, $i = 1, \ldots, m$, the integral in (6) becomes

$$V\{S_{\underline{0}}(y)\} = \tag{7}$$

$$\frac{y^m}{\prod_{i=1}^{m} w_i} \int_{z_1=0}^{1} \int_{z_2=0}^{1-\sum_{i=1}^{1} z_i} \ldots \int_{z_m=0}^{1-\sum_{i=1}^{m-1} z_i} dz_m \ldots dz_1.$$

The integral in (7) is the hyper-volume of the unit simplex

$$S = \{\, \underline{u} \mid \sum_{i=1}^{m} u_i \le 1, u_i \ge 0, i = 1, \ldots, m \,\}. \tag{8}$$

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

$$\frac{\Gamma(\eta)}{\prod_{i=1}^{m+1} \Gamma(\eta \cdot \nu_i)} \cdot \left(\prod_{i=1}^{m} (u_i)^{\eta \cdot \nu_i - 1} \right) \left(1 - \sum_{i=1}^{m} u_i \right)^{\eta \cdot \nu_{m+1}} \tag{9}$$

where $\eta > 0$, $\nu_i > 0$, $\sum_{i=1}^{m+1} \nu_i = 1$ has support S (cf. (8)) and by setting its parameters in (9) equal to $\eta = m+1$, $\nu_i = \frac{1}{m+1}$, $i = 1, \ldots, m+1$ it immediately follows from (8), (9) and the fact that $S_{\underline{0}}(y) = \emptyset$ for $y < 0$ that

$$V\{S_{\underline{0}}(y)\} = \frac{y^m}{m! \prod_{i=1}^{m} w_i} \cdot 1_{[0,\infty)}(y). \tag{10}$$

Again changing variables $x_i = u_i - v_i$, $i = 1, \ldots, m$ we arrive utilizing (4) at

$$V\{S_{\underline{v}}(y)\} = V\{S_{\underline{0}}(y - \sum_{i=1}^{n} w_i v_i)\}. \tag{11}$$

The lemma now follows from (10) and (11). □

***Theorem* 1 :** *The cdf of the weighted linear combination Y given by (1), where U_i are independent $[0,1]$ uniform random variables is given by (2).*

***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 2), 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 $Pr(Y \le y_1)$ for the value of y_1 indicated by Figure 2A and that of $Pr(Y \le y_2)$ for the value of y_2 depicted in Figure 2B. Figure 2A displays C^3 and $S_{\underline{0}}(y_1)$ (cf. (4)). Figure 2B 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 2A $Pr(Y \le y_1) = V\{S_{\underline{0}}(y_1)\}$. In Figure 2B the calculation of $Pr(Y \le y_2)$ is somewhat more complicated. Figure 2B shows that

$$Pr(Y \le y_2) = V\{S_{\underline{0}}(y_2)\} - V\Big\{ \bigcup_{i=1}^{3} S_{\underline{e}^i}(y_2) \Big\}. \tag{12}$$

Note that (12) also holds for the value y_1 in Figure 2A as $S_{\underline{e}^i}(y_1) = \emptyset$, $i = 1, \ldots 3$. Generalizing to $\mathbb{R}^m$ we obtain directly

$$Pr(Y \le y) = V\{S_{\underline{0}}(y)\} - V\Big\{ \bigcup_{i=1}^{m} S_{\underline{e}^i}(y) \Big\} \tag{13}$$

The inclusion-exclusion principle (cf. (3)) yields

$$V\Big\{ \bigcup_{i=1}^{m} S_{\underline{e}^i}(y) \Big\} = \sum_{i=1}^{m} V\{S_{\underline{e}^i}(y)\} - \tag{14}$$

$$\sum\sum_{i<j} V\{S_{\underline{e}^i}(y) \cap S_{\underline{e}^j}(y)\} +$$

$$\sum\sum\sum_{i<j<k} V\{S_{\underline{e}^i}(y) \cap S_{\underline{e}^j}(y) \cap S_{\underline{e}^k}(y)\} -$$

$$\ldots + (-1)^m V\Big\{ \bigcap_{i=1}^{m} S_{\underline{e}^i}(y) \Big\}.$$

Utilizing (4) it follows that the intersections of the simplices $S_{\underline{e}_i}(y)$ in (14) are all of following form

$$\bigcap_{i \in I} S_{\underline{e}^i}(y) = S_{\underline{v}}(y) \tag{15}$$

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 2B is the intersection of $S_{(1,0,0)}(y_2)$ and $S_{(0,1,0)}(y_2)$. From (15), (14) and (12) we conclude that

$$Pr(Y \le y) = \sum_{v_1=0}^{1} \ldots \sum_{v_m=0}^{1} (-1)^{\sum_{i=1}^{m} v_i} V\{S_{\underline{v}}(y)\}. \tag{16}$$

The proof of the theorem follows from Lemma 1. □

From the proof it follows that an efficient method to evaluate the distribution in (2) 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 (5) when a vertex is visited by the procedure. An example discussing the effect of dependence in the

Project Risk Analysis domain utilizing the cdf given by (2) is presented in the next section.

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

$$\begin{cases} E[T] = \dfrac{a+4m+b}{6} \\ Var[T] = \frac{1}{36}(b-a)^2. \end{cases} \tag{17}$$

Solving for the beta parameters using (17) has been controversial (see e.g. Clark (1962), Grubbs (1962)) and its use is still subject to a discussion (see, e.g. Kamburowski (1997)). Van Dorp & Kotz (2002) suggested the use of a Two-Sided Power (TSP) distribution, an extension of the triangular distribution, defined by the density

$$f_X(x|a,m,b,n) = \tag{18}$$
$$\begin{cases} \frac{n}{(b-a)}\left(\frac{x-a}{m-a}\right)^{n-1} & a < x \leq m \\ \frac{n}{(b-a)}\left(\frac{t-x}{b-m}\right)^{n-1} & m \leq x \leq b\,, \end{cases}$$

as a proxy to the beta, specifically in problems of assessment of risk and uncertainty (such as in PERT). For $n = 2$ in (18) the TSP density coincides with the density of a triangular distribution. The expressions for the mean and the variance for (18) result in

$$E[X] = \frac{a+(n-1)m+b}{n+1} \tag{19}$$

and

$$Var(X) = \tag{20}$$
$$(b-a)^2 \cdot \left\{ \frac{n - 2(n-1)\frac{(m-a)}{(b-a)}\frac{(b-m)}{(b-a)}}{(n+2)(n+1)^2} \right\}.$$

For a TSP distribution with $n = 5$, the mean values $E[T]$ in (17) and $E[X]$ in (19) coincide.

In the example to be discussed below, the effect of an assumption of independence between activity durations on the minimal completion time of a PERT network combined using the above setup (i.e. selecting either a beta, triangular or TSP distribution while utilizing the estimates a, m and b) will be compared to one associated with an assumption of dependence combined with triangular distributions.

3.1 *Description*

Figure 3 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 1. 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 Engineering Change Orders and crane unavailability.

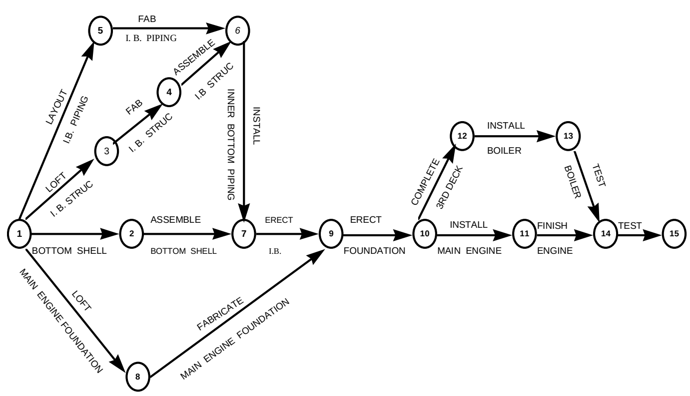

Figure 3. Project Network $\mathcal{P}$ for Production Process.

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 rank correlation between the activities and its associated aggregated risk are specified in Table 2. 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.

3.2 *Project Completion Time Distribution Analysis*

To show the effect of mild dependence between the activity durations on the minimal completion time distribution of the project in Figure 3, the information in Tables 1 and 2 and the dependence model described above have been used. A rank correlation of 0.5 is assumed across the board and may be viewed as a mild form of dependence.

Table 1. Parameters for modeling the uncertainty in activity durations for the project network in Figure 3.

ID	Activity Name	a	m	b
1	Shell: Loft	22	25	30
2	Shell: Assemble	35	37	43
3	I.B.Piping: Layout	19	22	29
4	I.B.Piping: Fab.	4	5	10
5	I.B.Structure: Layout	23	26	31
6	I.B.Structure: Fab.	16	18	24
7	I.B.Structure: Assemb.	11	14	20
8	I.B.Structure: Install	6	7	12
9	Mach Fdn. Loft	25	28	33
10	Mach Fdn. Fabricate	33	35	40
11	Erect I.B.	27	30	37
12	Erect Foundation	6	7	11
13	Complete #rd DK	4	5	9
14	Boiler:Install	6	7	10
15	Boiler:Test	9	10	15
16	Engine: Install	6	7	12
17	Engine: Finish	17	20	26
18	FINAL Test	13	15	20

Amongst the TSP and beta distribution, the triangular distribution is the only distribution that is completely specified by a, m and b without additional assumptions. Hence, the minimal completion time distribution involving triangular distributions and an assumption of mild dependence is compared with the project completion time distribution assuming independence between the activity durations with a triangular form (cf. (18) with $n = 2$), a beta form (via (17) and employing the method of moments) and finally a TSP form (cf. (18) with $n = 5$). The Monte Carlo analysis results utilizing 10,000 CPM calculations per case are displayed in Figure 4. For robustness purposes of the dependence model herein, Figure 4 also contains the minimal completion time distribution of the project utilizing triangular distributions for activitaty durations involving complete dependence (the strong dependence case in Figure 4). Complete dependence can be specified using the dependence model above by assigning rank correlations of 1 for all activities and mixture weights such that $w_{i^*j} = 1$ and $w_{ij} = 0$, $i \neq i^*$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ for some $i^* \in \{1, \ldots, m\}$. Finally, the minimal completion time of 126 days (CPM-Best Case), 144 days (CPM-Case) and 196 days (CPM-Worst Case) of a standard CPM analysis utilizing only the lower bounds, most likely estimates and upper bounds in Table 1, respectively, in are depicted by vertical lines.

Table 2. Parameters for modeling the dependence between activity durations for the project network in Figure 3.

ID	Activity Name	w_{ECO}	w_{CRANE}	ρ
1	Shell: Loft	1.0	0.0	0.5
2	Shell: Assemble	1.0	0.0	0.5
3	I.B.Piping: Layout	0.5	0.5	0.5
4	I.B.Piping: Fab.	1.0	0.0	0.5
5	I.B.Structure: Layout	1.0	0.0	0.5
6	I.B.Structure: Fab.	1.0	0.0	0.5
7	I.B.Structure: Assemb.	0.5	0.5	0.5
8	I.B.Structure: Install	0.5	0.5	0.5
9	Mach Fdn. Loft	0.5	0.5	0.5
10	Mach Fdn. Fabricate	0.5	0.5	0.5
11	Erect I.B.	0.2	0.8	0.5
12	Erect Foundation	0.2	0.8	0.5
13	Complete #rd DK	0.2	0.8	0.5
14	Boiler:Install	0.0	1.0	0.5
15	Boiler:Test	1.0	0.0	0.5
16	Engine: Install	0.0	1.0	0.5
17	Engine: Finish	1.0	0.0	0.5
18	FINAL Test	1.0	0.0	0.5

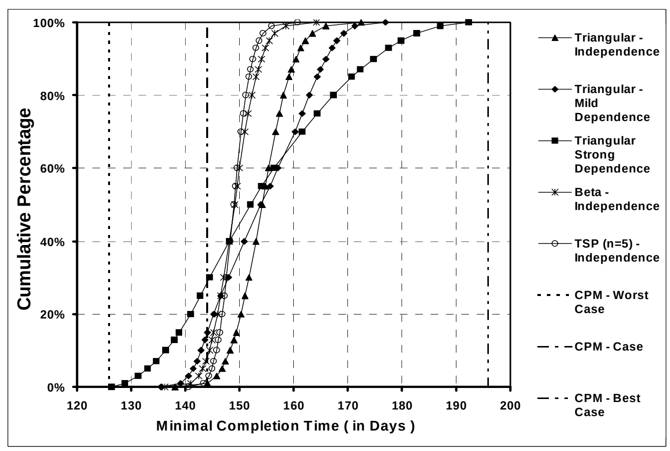

Figure 4. Comparison of Distributions of Minimal Completion Time for the project in Figure 3.

The minimum, mean, maximum, the standard deviation and range of the project completion distribution for the five combinations are provided in Table 3. Comparing the first and fourth row in Table 3 it follows that with the independence assumption between beta activity durations, the use of (17) 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 trian-gular distributions whose parameters are directly specified by the three estimates a, m and b (See Table 1). Hence, the adoption of (17) may not be consistent with a conservative approach towards estimating project completion time and its uncertainty. Note that from the fifth row in Table 3 follows that when utilizing TSP distributions ((18) 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.

Table 3. Minimum, Mean, Maximum, Standard Deviation and Range of the Project Completion Time Distribution using Triangular, Beta and TSP $(n = 5)$ Distributions under an Independence assumption and Triangular distributions under a Mild and Strong dependence assumption. Results were generated utilizing Monte Carlo Analysis involving 10,000 CPM calculations per case.

	Min	Mean	Max.	St. Dev.	Range
TRIANG - Independence	138.11	155.15	172.49	5.04	34.38
TRIANG - Mild Dependence	135.56	155.10	176.93	8.90	41.37
TRIANG - Strong Dependence	126.29	154.94	192.29	14.64	66.00
BETA - Independence	136.27	150.01	164.14	4.06	27.87
TSP (n=5) - Independence	140.63	149.85	160.78	2.96	20.15

The most notable results in Figure 4 and Table 3, however, follow from comparing the completion time distribution under an assumption of mild dependence and strong dependence (rank correlations equal to 1) with the distributions assuming independence. Although no mean shift occurs when comparing the first, second and third rows in Table 3, the standard deviation of the completion time of the project almost doubles (triples) when comparing the second (third) row to the first one. The same observation follows from Figure 4 where the distributions under the dependence assumptions posses much smaller slopes and appear to have a support that overlaps all of its counterparts. The latter observation is substantiated by comparing the minimum and maximum observed values for the minimal project completion time in the second and third row to those in the first, fourth and fifth row of Table 3. In addition, observe from Table 3 that the support of the distribution in the strong dependence case (third row) covers 94.29% of the possible range of 126 days (CPM - Best Case) to 196 days (CPM - Worst Case), whereas the accompanying independence case (first row) covers only 49.11%. The latter observation demonstrates the flexibility of the dependence model described in this paper, but also emphasizes the need for accurate assessment of the dependence parameter within the model.

Evidently, if the use of (17) 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)), the common assumption of independence between marginal distribution of activity durations should be subjected to the same level of scrutiny. Perhaps, such a level of scrutiny could lead to development of dependence models similar to the one described herein and the development of formal methods for assessing dependence parameters utilizing expert judgment elicitation.

Note also that it follows from Figure 4 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 mild 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, coupled with an independence assumption, could serve as an explanation for a low incidence of project success (on-time) when utilizing standard CPM analysis as a yard stick.

4 ACKNOWLEDGMENT

The author is indebted to Technical Program Board of the European Safety and Reliability Conference 2003 Conference and to the referees. Their comments significantly improved the content and presentation of the first version.

5 REFERENCES

Barrow, D.L. and 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*. New York: Oxford University Press.

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

Clemen, R.T. 1995. *Making Hard Decisions, An Introduction to Decision Analysis.* New York: Duxbury Press.

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

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.

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

Genest C. and 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.* Washington D.C.: The George Washington University (Masters Thesis).

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

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

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: 79-88. North Holland: Elsevier.

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.

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

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.

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.

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

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. New York: SNAME.

Van Dorp, J.R. & Duffey, M.R. 1999. Modeling Statistical Dependence in Risk Analysis for Project Networks. *International Journal of Production Economics* 58: 17-29.

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

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

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

Yi, W. and 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.

6 APPENDIX

The procedure $CalcCDF(\boldsymbol{G}, y, m, \underline{w})$ below evaluates the c.d.f. of Y given by (1). The algorithm uses functions

$ProdWeights(\underline{w}, m)$

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 $\psi = \sum_{i=1}^{m} w_i v_i$.

$VisitVertices(\boldsymbol{G}, y, i, \underline{v}, m, \underline{w}, \Pi);$
$Step\ 1: if\ i < m\ then\ v_i := 0;$
$visitVertices(\boldsymbol{G}, y, i, \underline{v}, sum, m, \underline{w}, \Pi);$
$v_i := 1;$
$VisitVertices(\boldsymbol{G}, y, i, \underline{v}, sum, m, \underline{w}, \Pi);$
$Step 2:$
$\Sigma := SumElements(\underline{v}, m);$
$\psi := SumProducts(\underline{v}, \underline{w}, m);$
$Step\, 3: If\ (y - \psi) > 0\, then$
$G := G + (-1)^{\Sigma} \frac{y-\psi}{m! \cdot \Pi}$

$CalcCDF(\boldsymbol{G}, y, m, \underline{w});$
$Step\, 1: If\ y \leq 0\ then\, G := 0; Stop;$
$Step\, 2: If\ y \geq 1\, then\, G := 1;\ Stop;$
$Step\, 3: \Pi := ProductWeights(\underline{w}, m);$
$Step\, 4: VisitVertices(\boldsymbol{G}, y, 1, \underline{v}, m, \underline{w}, \Pi);$