ELSEVIER

Int. J. Production Economics 58 (1999) 17–29

international journal of production economics

Statistical dependence in risk analysis for project networks using Monte Carlo methods

J.R. van Dorp, M.R. Duffey*

School of Engineering and Applied Science, The George Washington University, Washington, DC 20052 USA

Received 1 October 1997; accepted 3 December 1997

Abstract

Monte Carlo simulation of project networks is increasingly used by engineering firms to analyze schedule/cost risk for bidding purposes. However, one serious methodological flaw of most Monte Carlo simulations is the assumption of statistical independence of activity durations in the network. In this paper, a method is proposed to model and quantify positive dependence between uncertainty distributions of activities. This method inherits the theoretically sound foundations of the rank correlation method, but provides a less cumbersome method to elicit dependency information from project engineers. Details of the methodology are described along with an example of project risk analysis in a manufacturing domain (shipbuilding).

Keywords: Schedule/cost uncertainty; Rank correlation; Engineering judgment

1. Introduction

Risk analysis on project networks is defined here as the quantification of uncertainty in project schedule. When activity durations are deterministically known, the Critical Path Method (CPM) is a straightforward and well known method for quantifying project schedule. However, quantifying uncertainty in project schedule resulting from activity duration uncertainty in closed form is still difficult. As a result one resorts to Monte Carlo simulation of the project networks: realizations of the activity durations are drawn from their uncertainty distributions after which quantification of project schedule follows using the CPM method. By executing this step repeatedly, an arbitrarily large number of realizations of project completion time may be generated allowing inference of its uncertainty. In addition, cost uncertainty with time-dependent effects can also be quantified by secondary calculations in the project network.

In general, a risk analysis is desired for large projects that have not been executed before and therefore involve a lot of uncertainty in project schedule and project cost. While some hard historical data may be available, mainly past experience of project engineers with similar projects is used to

* Corresponding author. The George Washington University, Department of Engineering Management, 2130 H Street, N.W., Suite 632, Washington, DC 20052, USA. Tel.: + 1 202 994 7173; fax: + 1 202 994 4606; e-mail: duffey@seas. gwu.edu.

PII: S0925-5273(98)00081-4

populate the parameters of the risk analysis model. A project network $\mathscr{P}$ is generally represented as a tuple $(\mathscr{A}, \mathscr{N})$, where $\mathscr{A}$ is a set of activities and $\mathscr{N}$ is a set of events modeling the transitions between the activities. When assuming uncertainty of the activities in the set $\mathscr{A}$, probability distributions are used to model the individual uncertainty in the duration of each activity. In the discussion that follows, these distributions will be referred to as *marginal* distributions. The beta distribution has been proposed as a marginal distribution in the past for the Project Evaluation Review Technique [1]. Although the beta distribution is flexible, specification of its parameters may not be straightforward for project analysts. The triangular distribution has also proven popular among project analysts due to its simplicity [2]. Parameters of a triangular distribution are readily obtained through elicitation of a lower bound, upper bound and a most likely estimate from project engineers.

To assist decision-making during the design stage, Monte Carlo simulation for this type of network has begun to be used by some front-running companies in manufacturing, construction, power and other industries. Applications being explored include internal approval and/or external bidding for multi-million dollar projects [3]; comparison of design alternatives; assessment of schedule risk in supplier chains; and contract design for schedule-related penalty clauses and payment milestones. Typically in industry, these Monte Carlo simulations are conducted using risk analysis extensions to commercial CPM-based software which have only just become available in the past few years [4,5] along with faster/cheaper hardware which makes implementation viable. Other project network simulation packages have been available much longer [6,7], but without widespread commercial industry interest. While the increasing use of quantitative risk analysis for engineering projects is encouraging, it also necessitates a re-examination of some of the underlying methodological assumptions. This paper specifically addresses one serious methodological flaw in traditional Monte Carlo simulation of project networks: the assumption of statistical independence for individual activities which share risk factors in common with other activities in the set $\mathscr{A}$. A more general discussion of risk analysis methodology issues (including cost modeling) and a proof-of-concept software tool developed by the authors are described in Duffey et al. [8].

2. The statistical independence assumption

Given $(\mathscr{A}, \mathscr{N})$, the uncertainty in project schedule is completely defined when the uncertainty in the activity durations has been quantified. To do this, the multivariate distribution between uncertainties in the activity durations must be specified. It is *not uniquely* defined by just modeling the marginal distribution of the uncertainty in each activity duration, yet most available software packages fail to indicate this fact and elicit only the marginal distributions from the user to generate uncertainty distributions for project schedule. This is the assumption of statistical independence, i.e. that the marginal distributions of uncertainty for individual activities in the set $\mathscr{A}$ completely define the multivariate distribution for project schedule.

It is intuitively obvious that this assumption is highly suspect for many large engineering projects which involve multiple activities of a similar type and/or have different activity types which are influenced by common risk factors. As an example in a manufacturing domain, consider production activities being planned for an as-yet untested, innovative ship design [8]. Many hull modules might require very similar activities for structural fabrication, for which engineers can identify one or more common risks due to a new robotic welding technique, alignment problems during erection, or potential engineering change orders (ECOs) due to design problems. Alternately, there may be dissimilar activities which are collectively influenced by some external risk factor. An example would be risk of bad weather for painting, outfitting of piping and electrical systems and other activities scheduled under the "open sky" in the same time period. As will be demonstrated, failure to model such types of risk dependence during Monte Carlo simulation can result in the underestimation of total uncertainty in project schedule. Consider a single simulation run of a two-activity network with identical uncertainty distributions: values sampled independently from the two activities would be equally likely to occur anywhere in the uncertainty range: one may be high and the other low, as well as both low (or high). As a consequence, any cumulative effect of a common risk factor will be neglected, and may even cancel out completely. In small networks this effect might not be disastrous, but as the size of the network grows the underestimation will increase. As a practical consequence, management could decide to invest in a particular project due to a misrepresentation of the level of uncertainty in project schedule/cost.

The form of dependence described here, in which large values of one marginal distribution tend to be associated with large values of another marginal distribution, is known as *positive* dependence and has long been recognized [9]. It will be argued that this form of dependence may be most appropriate to model statistical dependence between the activity durations in project networks. Regardless of what form of dependence is chosen to model the statistical dependence, a complete multivariate distribution for the uncertainties in the activity durations needs to be specified. To do this, it is desirable to separate the modeling of the marginal distributions and the dependence effects. Marginal distributions are readily obtained using existing practice to elicit parameters for triangular or beta distributions from project engineers. A method to correlate dependency effects seperately is desirable.

Recent work in project risk analysis [2] specifies one method for modeling dependence between a small group of uncertainty distributions through *rank correlation methods* [10], allowing the marginal distributions to be specified separately. Rank correlations between uncertainty distributions were first introduced by Spearman [11]. The rank correlation measures a degree of positive dependence and is invariant under non-decreasing transformation of the uncertainty distributions. Kruskal [12] mentions that due to this invariance the rank correlation is an appropriate measure of positive dependence. However, specifying the degree of dependence through rank correlations using project engineers is difficult. Although theoretically sound, the measure is difficult to interpret even for project analysts with a strong background in probability and statistics. The other difficulty is sheer size: degree of dependence is typically specified by degree of correlation through a correlation matrix. A small network of 100 activities would ask for 10 000 correlations to be specified. Vose [2] recognizes these difficulties, and as an alternative he proposes the graphical *envelope method*. In the envelope method historical data regarding two activity durations is plotted in a scatter plot and two bounding lines (the envelopes) are drawn such that all data points are contained within the lines. When a realization is sampled for the one activity duration, the sample of the other activity duration is sampled between the bounding lines. However, even though the envelope method might work for a small number of uncertainty distributions, it lacks the theoretical foundation of rank correlation methods and does not seem to offer a structured approach for dealing with large sets of dependent uncertainty distributions.

To summarize, the importance of relaxing the independence assumption has been clearly recognized by authorities on project risk analysis methodology [2,13]. However, techniques to model statistical dependence have not yet been incorporated into most of the popular commercial software packages for Monte Carlo simulation of project networks. The reasons for this may be two-fold. First, while project analysts increasingly model uncertainties in activity durations using distributions like the beta and the triangular, many project decision-makers remain generally skeptical of quantitative methods. This skepticism would even be higher when dealing with an issue more abstract than uncertainty (statistical dependence). Second, having convinced a project analyst to model dependence, the degree of dependence needs to be specified. Due to the difficulties in existing methods described above, even highly motivated project analysts would be discouraged by such a task and resort to assuming statistical independence.

To address the limitations described above, a methodology is described below which: (1) offers a structured approach for modeling dependence between large sets of uncertainty distribution, (2) allows the marginal distributions to be specified separately, (3) inherits the theoretical foundations of the rank correlation methods, (4) asks for a dependence measure which can be interpreted by project engineers (5) asks a number of dependence measures to be specified that is only as large as or smaller than the total number of activities in the project network and finally (6) samples efficiently from the resulting multivariate distribution.

3. A statistical dependence model for project networks

To model the statistical dependence between the uncertainties in the activity durations a multivariate distribution of the uncertainties in the activity durations need to be modeled. Assuming that the marginal distributions are to be specified separately, the two extremes in this case are: (1) assuming statistical independence between the marginal distributions and (2) specifying a complete multivariate distribution exhibiting dependence with the pre-specified marginal distributions as its marginals. In the previous section, it was argued that the first extreme may be specious. The second extreme is generally achieved by specifying a multivariate distribution of a known family (e.g. the multivariate normal family) with a particular rank correlation matrix. This known distribution may be easily transformed in a multivariate distribution on a unit hypercube having uniform distributions on [0, 1] as its marginals. Using standard distribution theory, any absolute continuous marginal distribution may be derived from the uniform marginal by an appropriate transformation. As the rank correlation is invariant under non-decreasing transformations, the rank correlation matrix of the transformed multivariate distribution is the same as the one initially specified. Note that this assertion is not true for the correlation matrix (as opposed to *rank* correlation matrix) of the initially specified multivariate distributions. Pathological examples exist where initially the correlation between two random variables is 1, but after transformation the correlation between the random variables is 0. The second extreme has the disadvantage that even for a small number of activities (e.g. 100) a rank correlation matrix has to be specified with 10 000 elements. Even if project engineers would be able to estimate such a matrix, the estimated matrix has to be a positive-definite matrix. In general, the estimated matrix turns out to be not positive definite and a modification method has to be specified to modify the estimated rank correlation matrix into the closest positive-definite matrix. All in all, the second extreme, though theoretically sound, seems not practical enough to be used by project analysts.

Between the extremes cited above, an intermediate method can be considered which: (1) assumes independence between marginal distributions by default but (2) allows joint distributions for subsets of uncertainty distributions which share common risk factors. Such a method was developed for modeling dependencies between uncertainty distributions in Van Dorp [14]. This method was implemented in the prototype software package UNICORN [23] and was used to estimate damage the Dutch Dike System due to the flooding in Spring 1995. The resulting multivariate distribution exhibits a known form of *positive dependence*. An additional advantage of a multivariate distribution exhibiting a known form of positive dependence is that for these types of distributions, a rank correlation of 0 does imply statistical independence. The multivariate distribution introduced in Van Dorp [14] is constructed in two steps. First, assumptions of independence need to be specified and second the joint distribution needs to be specified between the remaining dependent random variables.

3.1. *Assumptions of conditional independence given a common risk factor*

The assumptions of independence in Van Dorp [14] uses the idea of latent variable models [15]. Latent variable models and factor analysis have found wide application in the behavioral sciences [16]. A classical example of a latent variable model is the following; consider a person performing two different types of valid IQ tests. Clearly, depending on the intelligence of the person the results of the IQ tests would be both high or both low. Prior to the execution of the tests the outcome of the tests may be modeled as random variables which are clearly positive dependent. The source of positive dependence is the unknown IQ level of the person executing the test. In a latent variable model, the IQ level would be identified as the latent random variable. The crucial assumption in Latent variable models is that were we to know the IQ level exactly prior to the execution of the intelligence test, the outcome of the tests are assumed independent. This latter assumption is known as the assumption of *conditional independence*.

The idea within latent variables models that unknown variables are the source for the statistical dependence can be applied to project risk analysis. A more appropriate name for such latent variables would be *common risk factors*. Take for example, the painting of four equal sized outside walls of a house by four different painters. Each painting of a wall may be identified as a separate activity. Assume the duration of each activity is heavily affected by the weather situation at the time of painting. The unknown weather situation can be identified as the *common risk factor* between the four activities. It seems reasonable to assume that, were we to know the exact weather situation at the time of painting that the uncertainties in the duration of each activity are independent. The idea of *common risk factors* or *common causes* is not new and has already found wide appreciation in risk analyses methods like fault tree analysis for chemical and nuclear power plants [17,18]. In fact, it is the predominant reason to apply the technique of diversification when designing redundancy in such plants for safety reasons. Brainstorming sessions are generally used in such applications to identify the common causes. Accordingly, brainstorming sessions involving project engineers may be used to identify the *common risk factors* for a particular project. Let the set of common risk factors be denoted by $\mathscr{F}$ and each individual risk factor be denote by $\mathscr{F}_i$. Next, the set of activities $\mathscr{A}$ needs to be divided in the brainstorming session into disjoint subsets $\mathscr{A}_i$ such that each subset $\mathscr{A}_i$ is predominantly affected by the risk factor $\mathscr{F}_i$. The subset of activities $\mathscr{A}_i$ may be identified as the *risk group* with *common risk factor* $\mathscr{F}_i$.

The assumptions of independence are formalized in the method in Van Dorp [14] through dependence diagrams. A dependence diagram is a directed graph $(\mathscr{R}, \mathscr{D})$ where $\mathscr{R}$ is a set of random variables and $\mathscr{D}$ is a set of arcs defining the dependence between the random variables. Let $\mathscr{R}$ be $\{r_1, \dots, r_n\}$. Let (r_i, r_j) be an arc in $\mathscr{D}$. Then r_i is an immediate predecessor of r_j and r_j is an immediate successor of r_i. Let $P(r_j)$ be the set of immediate predecessors of r_j, $j = 1, \dots, n$ and $S(r_i)$ be the set of immediate successors of r_i, $i = 1, \dots, n$. Let $R(\mathscr{R}, \mathscr{D})$ be the set of random variables such that $|P(r_j)| = 0$. That is, $R(\mathscr{R}, \mathscr{D})$ is the set of root nodes of the directed graph.

Definition. A dependence diagram is a directed graph $(\mathscr{R}, \mathscr{D})$ *such that*

- Nodes represent random variables while directed arcs indicate positive or negative dependence between end nodes of the arc.
- The random variables $r_i \in R(\mathscr{R}, \mathscr{D})$ are independent random variables.
- $\forall r_i \in \mathscr{R}$: $|P(r_j)| \in \{0, 1\}$.
- $\forall r_i \in \mathscr{R}$:$r_j \in S[r_i]$ are independent given the state of the immediate predecessor.

Typically when using dependence diagrams in project risk analysis, the set of root nodes $R(\mathscr{R}, D)$ would be the set of common risk factors $\mathscr{F}$. Fig. 1 contains an example of such a dependence diagram.

In Fig. 1, activity durations in the risk group consisting of A_1, A_2 and A_3 are independent given the state of the common risk factor $\mathscr{F}_1$. The risk group consisting of A_4, A_5 and A_6 is independent given the state of the common risk factor $\mathscr{F}_2$. The state of the common risk factors $\mathscr{F}_1$ and $\mathscr{F}_2$ are assumed independent.

As one practical example, consider a situation in the shipbuilding domain. A_1, A_2 and A_3 could be uncertainties in activity durations required to paint three separate modules of a ship in the same time period and $\mathscr{F}_1$ could be the severity of inclement weather during those days. Were we to know the weather situation at the time of painting a reasonable assumption would be that A_1, A_2 and A_3 are independent. A_4, A_5 and A_6 could be the uncertainties in the activity durations concerning the fabrication and welding of these three modules that need to be painted. $\mathscr{F}_2$ could be the magnitude of Engineering Change Orders (ECO) due to difficulty of accuracy control with respect to these modules. Again, were we to know the amount of ECOs in

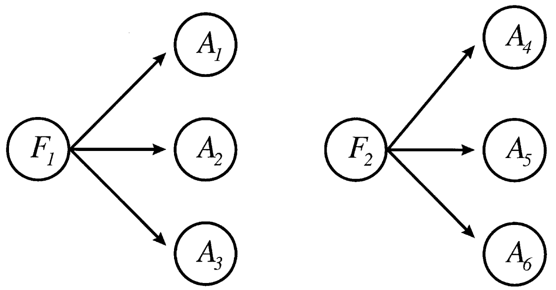

Fig. 1. An example dependence diagram.

advance, a reasonable assumption would be that A_4, A_5 and A_6 are independent.

3.2. Multivariate distributions between subsets of uncertainty distributions

Consider the example in Fig. 1. Ultimately, a multivariate distribution needs to be specified between A_1, A_2, A_3, A_4, A_5 and A_6. The marginal distributions of A_1, A_2, A_3, A_4, A_5 and A_6 are available and assumptions of conditional independence were specified by introducing new random variables, the common risk factors $\mathscr{F}_1$ and $\mathscr{F}_2$. Through the assumptions of independence, joint distributions need to be specified between A_1, A_2, A_3 and between A_4, A_5 and A_6.

If desired one may integrate over the common risk factor $\mathscr{F}_1$ to obtain the joint distribution between A_1, A_2, A_3. There is one crucial advantage to following this approach. Rather than having to ask a project engineer about the joint distribution between uncertainties in the activity durations A_1, A_2, A_3, the project engineers, through the assumption of conditional independence, can be asked about the joint distribution between $\mathscr{F}_1$ and A_i, $i = 1, \ldots, 3$. For the problem at hand with only three variables this may not seem such a big advantage. In practical applications, however, it is quite possible that 10 or more uncertainty distributions are affected by the same common risk factor. Two disadvantages of following this approach are: (1) to specify a joint distribution between $\mathscr{F}_1$ and A_i, one would have to ask about the marginal distribution of the common risk factor $\mathscr{F}_1$ and (2) to obtain the joint distribution between A_1, A_2, A_3 one would have to integrate over the random variable $\mathscr{F}_1$. As it turns out both disadvantages may be eliminated.

Consider first specifying the joint distribution (and thereby statistical dependence) between $\mathscr{F}_1$ and A_1 (a bivariate distribution). Note that the marginal distributions for all A_i, $i = 1, \ldots, n$, are already specified by project engineers. A natural modeling method to model bivariate distributions with specified marginal distributions is the method of Copula's [19]. The joint distribution between $\mathscr{F}_1$ and $\mathscr{A}_1$ with specified marginals for $\mathscr{F}_1$ and $\mathscr{A}_1$ is uniquely determined by its associated Copula. A Copula is a bivariate distribution with its marginals uniform distributions on $[0, 1]$. Sampling from the joint distribution of $\mathscr{F}_1$ and A_1 through Copulas is straightforward. First, one samples a bivariate sample from the Copula and next transforms the two realizations of the uniform variates into realizations of $\mathscr{F}_1$ and $\mathscr{A}_1$ through the inverses of their cumulative distribution functions. An additional advantage of using a Copula is that to create a sample of A_1, it is not necessary to know the marginal distribution of $\mathscr{F}_1$. Instead, one could sample a bivariate sample from the Copula and only transform the uniform variate associated with A_1 into a realization of A_1. Note that the first disadvantage of specifying the joint distribution between A_1, A_2, A_3 through the common risk factor $\mathscr{F}_1$ is eliminated. The elimination of the second disadvantage of using the common risk factor $\mathscr{F}_1$ is addressed in Section 3.3.

There are numerous choices of families of one-parameter Copulas that may be used to define the bivariate distribution between $\mathscr{F}_1$ and A_1. Generally, after the correlation is defined between its two uniform marginals, only a single parameter of the Copula needs to be specified. As it turns out, the correlation between the uniform marginals of the Copula is exactly the rank correlation between $\mathscr{F}_1$ and A_1. It was already argued that asking for a rank correlation from project engineers is impractical as rank correlations are difficult to interpret. The Copula we propose to use is the Diagonal Band distribution first introduced by Cooke and Waij [20]. We argue that the advantage of using the Diagonal Band distribution is that: (1) it is efficient to sample a bivariate sample from the

Diagonal Band distribution and (2) the parameter of the Diagonal Band distribution may be elicited from project engineers through a quantity which can be more easily interpreted. In addition, in Van Dorp [14] it was shown that the driving factors in two different uncertainty analyses projects using other Copula's as well as the Diagonal Band distribution, were the marginal distributions for $\mathscr{F}_1$ and A_1 and the rank correlation between $\mathscr{F}_1$ and A_1, i.e. not the choice of Copula modeling the bivariate distribution.

Fig. 2 gives an example of a bivariate Diagonal Band distribution $D(U, V)$ of two uniform on [0, 1] distributed random variables U and V. Let the marginal cumulative distribution function of $\mathscr{F}_1$ be denoted by $\mathscr{H}$ and the marginal cumulative distribution function of A_1 be denoted by $\mathscr{G}$. It is well known that $\mathscr{H}(\mathscr{F}_1)$, $\mathscr{G}(\mathscr{A}_1)$ are uniform random variates on [0, 1]. Hence, in Fig. 2 U may be associated with $\mathscr{H}(\mathscr{F}_1)$ and V may be associated with $\mathscr{G}(A_1)$. Let the parameter of the diagonal band distribution be denoted by θ. Fig. 2 also relates the diagonal band distribution to its parameter θ.

The probability density $d(u, v)$ in Figs. 2 and 3 is distributed as follows:

$$d(u,v) = \begin{cases} \dfrac{1}{(1-\theta)} & \textit{Area } 1, \textit{Area } 5 \text{ (end regions of diagonal band)}, \\ 0 & \textit{Area } 2, \textit{Area } 4 \text{ (i.e., outside diagonal band)}, \\ \dfrac{1}{2(1-\theta)} & \textit{Area } 3 \text{ (rectangular region of diagonal band)}. \end{cases} \quad (1)$$

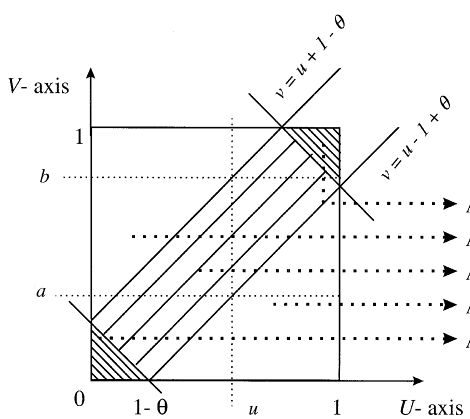

Fig. 2. A diagonal band distribution seen from above.

Note that (1) $\theta = 0$ implies U and V are independent, (2) $\theta = 1$ implies U and V are identical and (3) $0 < \theta < 1$ specifies an intermediate degree of positive dependence.

It may be derived that given the value u in Fig. 2, the conditional distribution $D(v|u)$ is a uniform variate on the interval $[a, b]$. For values of u closer to 0 and to 1 the conditional distribution is slightly more complicated, but can be easily obtained. Sampling a bivariate sample from a diagonal band distribution with parameter θ is straightforward. The method is described below in pseudo Pascal. First a sample u from a uniform random variate on [0, 1] needs to be sampled. The associated value of v is then sampled as follows:

Step 1: Sample v from a Uniform Random Variate on [0, 1].
Step 2: $a := u - 1 + \theta$; $b := u + 1 - \theta$. (2)
Step 3: $v := (b - a) \cdot v + a$.
Step 4: If $v < 0$ then $v := -v$;
If $v > 1$ then $v := 1 - v$.

The next section discusses a method to elicit the dependence parameter θ of the diagonal band distribution.

3.3. Eliciting the degree of dependence from project engineers

The idea behind the dependence measure that may be asked from project engineers lies in the comparison of the conditional distribution $D(v|u)$ and the marginal distribution $D(v)$, which is a uniform random variate on [0, 1]. The uncertainty in $D(v)$, and indirectly the uncertainty in $\mathscr{A}_1$, ranges from 0 to 1 and covers 100% of the range of V, and thus 100% of the range of A_1. For the case depicted in Fig. 2, the uncertainty in $D(v|u)$ that remains, ranges from a to b which covers $(b - a)$% of the range of V. It is argued that the above may be interpreted such that $100 - (b - a)$% of the uncertainty in V, and therefore in $\mathscr{A}_1$ is explained by knowing the value of u, i.e. knowing the value of the common risk factor $\mathscr{F}_1$. Again for values of u closer to 0 and 1 the analysis is slightly different but may be derived analogously. Averaging the % explanation of U in V, and therefore the %

explanation of $\mathscr{F}_1$ in $\mathscr{A}_1$, over all possible values of u, yields

$$\varepsilon = \theta^2 100\% \tag{3}$$

where ε is the average % explanation of the common risk factor $\mathscr{F}_1$ in the uncertainty of the activity duration $\mathscr{A}_1$. Note, that when $\theta = 1$, the average % explanation is 100% which coincides with the case of the diagonal band distribution where U and V are identical. When $\theta = 0$, the average % explanation is 0%, which coincides with the diagonal band distribution where U and V are independent. A possible question to ask from project engineers to solve for the parameter θ in the case of the common risk factor $\mathscr{F}_1$ and uncertainty in activity duration $\mathscr{A}_1$ would be

> Suppose you were to know the exact weather situation at the time of painting the of the ship modules, what percentage of your original uncertainty in the duration is explained?

If the project engineer answers 0%, this means that the weather has no influence on the uncertainty in the duration. If the project engineer answers 100%, this means that all his uncertainty in the duration is explained by not knowing the weather at the time of painting. If the project engineer answers X%, the parameter of the diagonal band distribution follows as

$$\theta = \sqrt{\frac{X}{100}}. \tag{4}$$

The interpretation of average % explanation by project engineers can be further improved by developing an elicitation process.

Although this average % explanation ε may still be less than intuitive for project engineers, an elicitation procedure can be developed which calculates ε indirectly by querying the expert about observable events (e.g., If the weather was X, what would the uncertainty distribution be?). Even as is, it is argued that ε is easier to interpret than asking for the rank correlation between the common risk factor $\mathscr{F}_1$ and the uncertainty in the activity duration $\mathscr{A}_1$. For illustration purposes Fig. 3 gives the relation between the average % explanation ε, the dependence parameter θ, and the rank correlation $\rho(\mathscr{F}_1, \mathscr{A}_1)$. It can be derived that for a diagonal band distribution with parameter θ as the Copula associated with $\mathscr{F}_1$ and $\mathscr{A}_1$, the rank correlation $\rho(\mathscr{F}_1, \mathscr{A}_1)$ equals

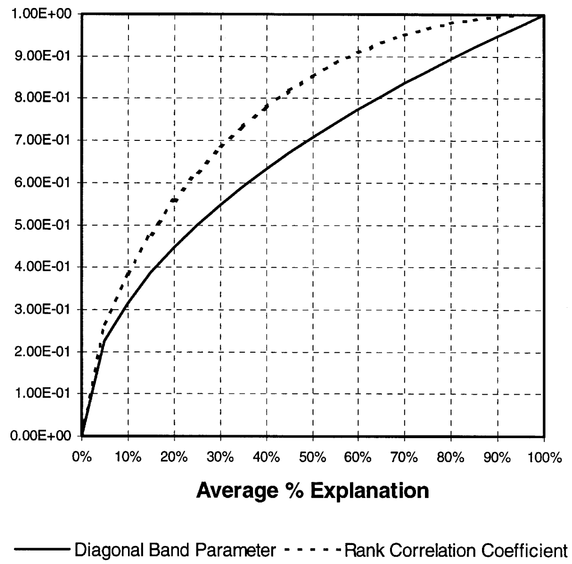

Fig. 3. Relations between the average % explanation, the diagonal band parameter and the rank correlation.

$$\rho(\mathscr{F}_1, \mathscr{A}_1) = -\theta^3 + \theta^2 + \theta. \tag{5}$$

Returning to the example dependence diagram in Fig. 1, the question above would have to be asked for every uncertainty in activity duration A_i, $i = 1, \ldots, 3$, with respect to the common risk factor $\mathscr{F}_1$. The parameter for each diagonal band distribution may be easily solved using Eq. (4). Finally, a sample from the joint distribution of $\mathscr{A}_i$, $i = 1, \ldots, 3$, may be easily obtained using the relation to the common risk factor $\mathscr{F}_1$. The procedure to create such a sample is described below. A sample from the joint distribution of A_4, A_5 and A_6 may be obtained analogously. Note that, a sample from the joint distribution of $\mathscr{A}_i$, $i = 1, \ldots, 3$, is generated using the procedure above without integrating the joint distribution of $\mathscr{A}_i$, $i = 1, \ldots, 3$, and $\mathscr{F}_1$ over the random risk factor $\mathscr{F}_1$. Hence, the second disadvantage of using the *common risk factor approach* is eliminated.

Step 1: Sample u from a Uniform Random Variate.
Step 2: Sample v_i from every diagonal band distribution associated with $\mathscr{F}_1$ and $\mathscr{A}_i$ using Eq. (2).
Step 3: Transform v_i into a realization of $\mathscr{A}_i$ using the inverse of the cummulative distribution function of $\mathscr{A}_i$. (6)

Following the approach presented here, (1) a structured approach which uses risk factors to model statistical dependence is offered, (2) the marginal distributions of the uncertainties in activity durations can be specified separately, (3) the theoretical foundation of the rank correlation methods as the parameter ε can be easily converted in a rank correlation, (4) a dependence parameter is asked for from project engineers which may be better interpreted than rank correlations, (5) the number of parameters to specify the statistical dependence is the same or less than the number of activity durations and (6) sampling from the resulting multivariate distribution is done in a computationally efficient manner.

4. Example project risk analysis

To demonstrate application of the above method to project network simulation, consider the small, 18-activity project network for a production process in Fig. 4 (from a well known text on ship design and construction [21]). Modern-day ship production is a manufacturing domain in which innovative design and build strategies require particular attention to risk factors that may impact cost and delivery time. One major risk area is the impact of engineering change orders (ECOs) and rework. Engineering changes can come from a variety of sources – owner-requested changes, inadequate design specifications, interface problems for vendor-furnished equipment, etc.

To show the effect of dependence between the activity durations, the dependence diagram given in Fig. 5 has been used to evaluate the minimum completion time of project $\mathscr{P}$. To model the uncertainty in each activity duration individually, the triangular distribution has been used given the parameters in Table 1. The sole risk source was

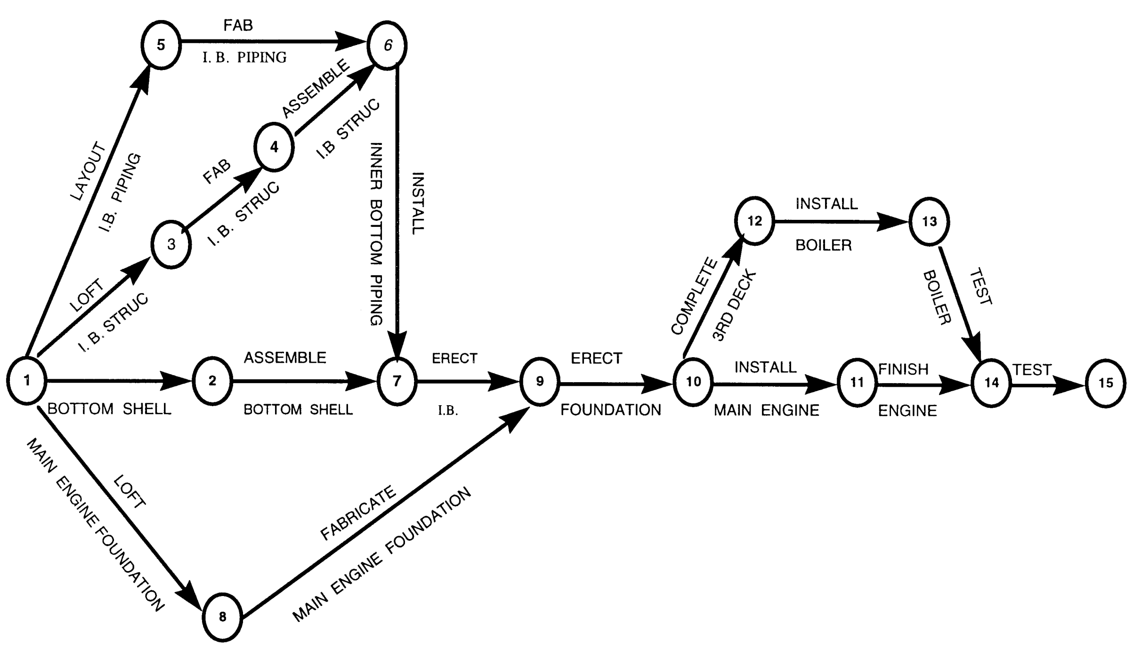

Fig. 4. Example project network $\mathscr{P}$ for production process.

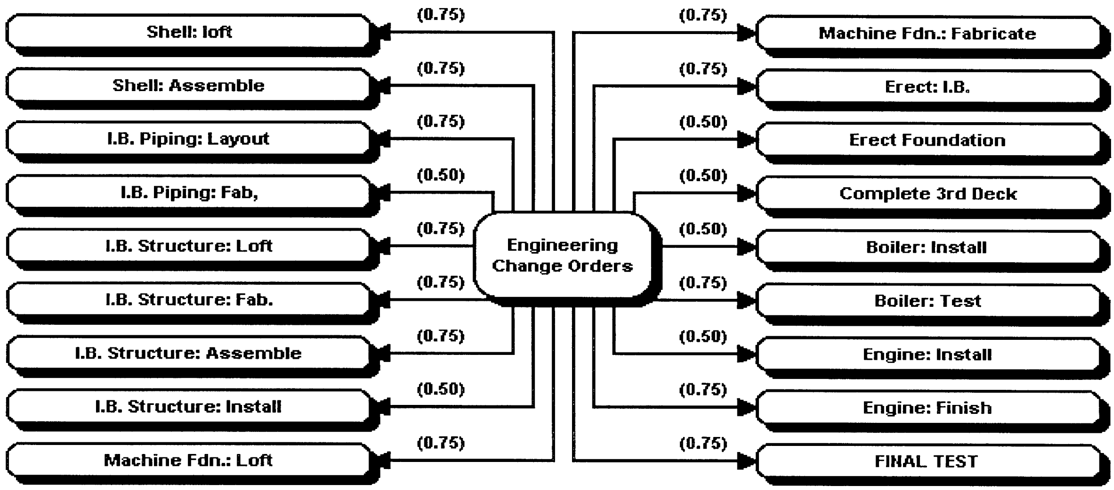

Fig. 5. Dependence diagram used in project $\mathscr{P}$ to model dependence.

Table 1
Standard PERT analysis of project $\mathscr{P}$

Activity	Lower bound	Most likely	Upper bound
Shell: loft	21	25	29
Shell: assemble	33	37	41
I.B. piping: layout	18	22	26
I.B. piping: fab	3	5	7
I.B. structure: loft	22	26	30
I.B. structure: fab.	14	18	22
I.B. structure: Assemb.	10	14	18
I.B. structure: install	5	7	9
Mach. Fdn.: loft	24	28	32
Mach. Fdn.: fabricate	31	35	39
Erect I.B.	26	30	34
Erect foundation	5	7	9
Complete 3rd DK	3	5	7
Boiler: install	5	7	9
Boiler: test	8	10	12
Engine: install	5	7	9
Engine: finish	16	20	24
Final test	11	15	19
Minimal completion time	117	144	174

assumed to be Engineering Change Orders. The number on the arcs is the dependence parameter ε between the magnitude of Engineering Change Orders (ECOs) and the uncertainty in the activity durations. Activities with an uncertainty range with less than 10 days were assigned an average % of explanation of 50%. Activities with an uncertainty range with more than 10 days were assigned an average % of explanation of 75%. The associated parameters of the diagonal band distributions may be easily solved for using Eq. (3). Next, rank correlations may be calculated using Eq. (4).

Fig. 6 gives the result in the uncertainty in the minimal completion time using dependence between the uncertainties in the activity durations. In addition, the optimistic, most likely and pessimistic estimates of a Standard PERT approach are shown as vertical lines in the plot.

For comparison, Fig. 7 gives the results using the independence assumption between the uncertainties in the activity durations. As can be seen from Figs. 6 and 7, the mean time for completion time is not very different in the two simulations of project network $\mathscr{P}$. However, the shape and bounds of the histograms have changed significantly. Assuming dependence between activity durations as given in Fig. 5, Fig. 6 indicates project completion with 95% certainty within *less than* 159.3 days as opposed to the 151.2 days in the case of assuming independence (a 5.4% increase in the time estimate).

5. Discussions and conclusions

In the example cited above, one might argue that the effect of dependence is exaggerated by the high

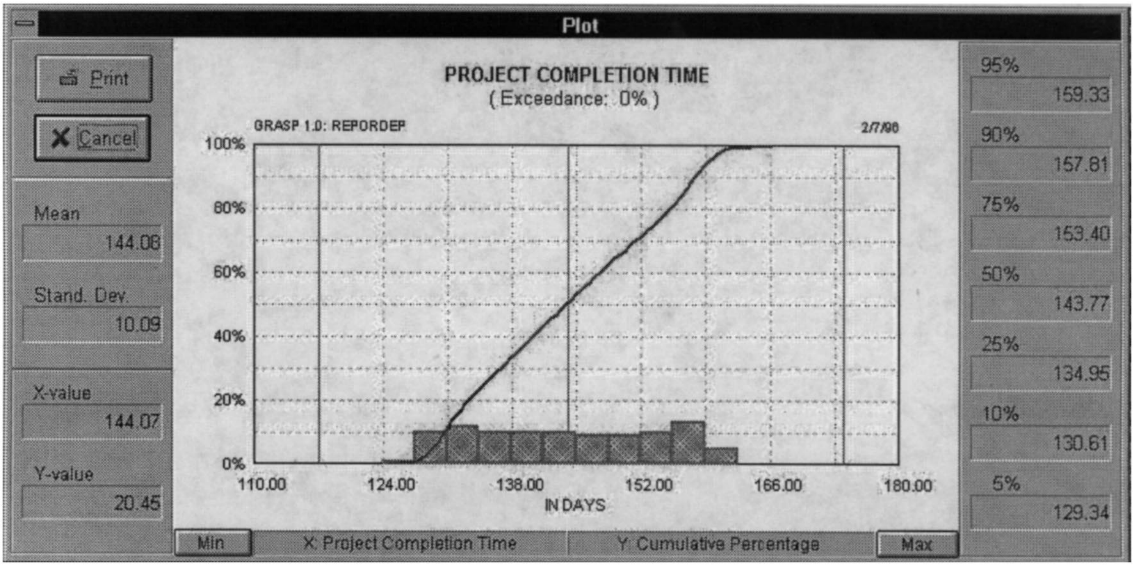

Fig. 6. Plot for uncertainty in minimal completion time of $\mathscr{P}$ with dependence.

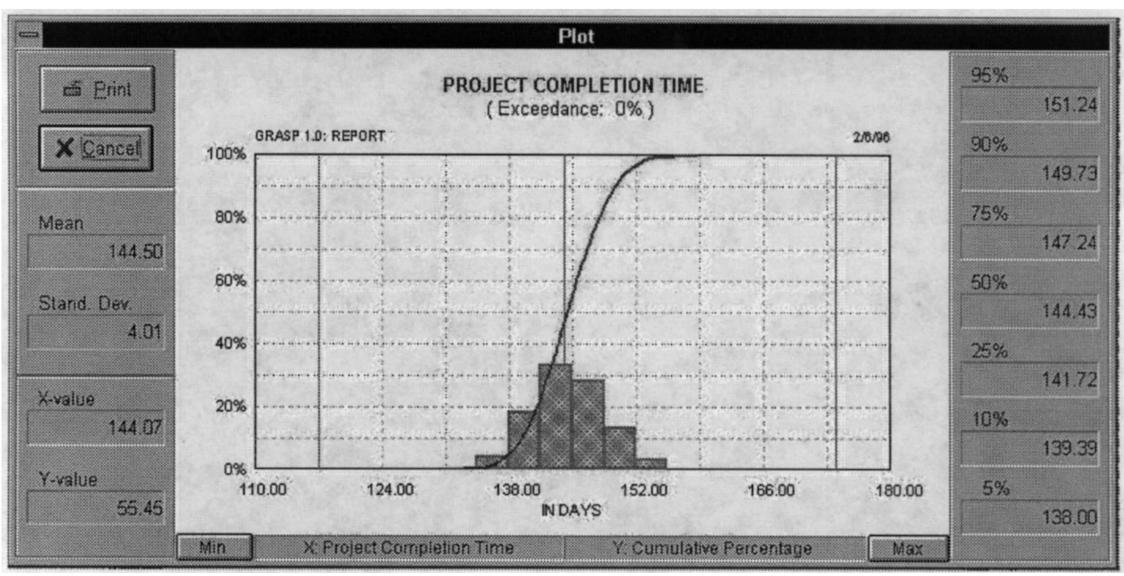

Fig. 7. Plot with independence assumption for uncertainty in minimal completion Time of $\mathscr{P}$.

values of the dependence parameter θ assumed in Fig. 5. However, our experience so far indicates that the effect of neglecting dependence will be more significant as the size of the project network grows, even with lower degrees of dependence assumed between the activity durations. For a second case study of a 250-activity shipbuilding network (typical in size of networks created by shipyards in the pre-contract bidding phase) and multiple dependency diagrams related to different activity classifications, the 95% certainty level for minimum completion time was 420.8 days with dependence vs. 383.7 days without dependence (a 9.7% increase). It is nonetheless difficult to generalize about scaling effects for networks of different complexities and different sets of dependency diagrams.

Clearly, however, the definition of valid risk groups for a specific engineering project would require careful consideration by project analysts for factors such as engineering change orders, subcontractor efficiencies, new production technologies, etc. In addition, there are other issues which need to be addressed to advance the methodology of simulation-based project risk analysis beyond existing commercial software. Such issues include better elicitation methods for uncertainty distributions from project experts; incorporation of activity-based costing models; learning curves effects; and advanced applications for design-stage decision-making. These other issues were addressed during implementation of proof-of-concept software for this project, and are discussed in a companion paper by Duffey and van Dorp [22]. Given the prevalence of risk-related production schedule delays experienced in many large, innovative engineering projects, use of dependence modeling might help provide more realistic estimation under uncertainty. The methodology also may be useful for other computer-based applications which use Monte Carlo simulation of activity networks, such as in business process reengineering.

Acknowledgements

We would like to thank several people for their advice and consultation on this project: Kevin Lyons at the National Institute of Standards and Technology (NIST); Philip Koenig and Marc Greenberg at the Naval Surface Warfare Center, Carderock Division; and Amita Singh (graduate research assistant, George Washington University). Financial support for this research has been provided in part by NIST.

References

[1] PERT, Summary Report, Phase 1, Special Projects Office, Bureau of Ordnance, Department of the Navy, Washington, DC, 1958.
[2] D. Vose, Quantitative Risk Analysis, A Guide to Monte Carlo Simulation Modeling, Wiley, New York, 1996.
[3] D. Roeleven, Building on Uncertainty, Collected Papers of the Workshop on Modeling the Product Realization Process: Innovative Methods, Computer Tools and Applications, Washington, DC, 1996.
[4] Primavera Systems, Inc., Monte Carlo for Primavera: Project Risk Analysis Software, Reference manual publication and web site information at www.primavera.com/products/, 1995.
[5] Welcome Software Technology, Guide to OPERA (Open Plan Extention for Risk Analysis), http://www.wst.com/products/opera.html, 1995.
[6] PAN, User Guide, Metier Management Systems, Ltd., London, 1981.
[7] H.M. Cockerham, RISNET Analyst's Guide, Huntsville, AL, 1972.
[8] M.R. Duffey, P. Koenig, D.W. Rosen, P. Singh, Design infrastructure in shipbuilding and other heavy industries, Transactions, Society of Naval Architects and Marine Engineers (SNAME) 104 (1996) 191–217.
[9] E.L. Lehmann, Some concepts of dependence, Annals of Mathematical Statistics 43 (1966) 1137–1153.
[10] R.L. Iman, W.J. Conover, A distribution-free approach to inducing rank order correlation among input variables, Communications in Statistics-Simulation and Computation 11 (3) (1982) 311–334.
[11] C. Spearman, The proof and measurement of association between two things, American Journal of Psychology 15 (1904) 2–101.
[12] H.K. Kruskal, Ordinal Measures of Association, Journal of American Statistical Association 53 (1958) 814–861.
[13] D. Cooper, C. Chapman, Risk Analysis for Large Projects, Wiley, New York, 1987.
[14] J.R. Van Dorp, Dependence Modeling for Uncertainty Analysis, Internal Report, Delft University of Technology, Delft, Holland, 1991.
[15] P.W. Holland, P.R. Rosenbaum, Conditional association and unidimensionality in monotone latent variable models, The Annals of Statistics 14 (4) (1986) 1523–1543.
[16] D.J. Bartholomew, Latent Variable Models and Factor Analysis, Oxford University Press, New York, 1987.
[17] D.F. Haasl, N.H. Roberts, W.E. Vesely, F.F. Goldberg, Fault Tree Handbook, Systems and Reliability Research, Office of Nuclear Regulatory Commission, Washington, DC, 1981.
[18] Q. Zhang, A general method dealing with correlations in uncertainty propagation in fault trees, Reliability Engineering and System Safety 26 (1989) 231–247.

[19] C. Genest, J. Mackay, The Joy of Copulas, bivariate distributions with uniform marginals, The American Statistician 40 (4) (1986) 280–283.
[20] R.M. Cooke, R. Waij, Monte Carlo sampling for generalized knowledge dependence, with application to human reliability, Risk Analysis 6 (3) (1986) 335–343.
[21] R. Taggart, Ship design and construction, The Society of Naval Architects and Marine Engineers (SNAME), New York, 1980.
[22] M.R. Duffey, J.R. van Dorp, Risk Analysis for large engineering projects: modeling cost uncertainty for ship production activities, Journal of Economic Valuation and Cost Analysis, (1998) forthcoming.
[23] R.M. Cooke, UNICORN, Methods and Code for Uncertainty Analysis, Atomic Energy Authority for the European Safety and Reliability Association, Delft, The Netherlands, 1995.