Monitoring Uncertainty in Project Completion Times: A Bayesian Network Approach

Ifechukwu C. Nduka[1], J. René van Dorp[2,*]

[1,2] Department of Engineering Management and Systems Engineering, The George Washington University, 800 22nd Street NW, Suite 2800, Washington, DC 20052, USA

Abstract

Recent advances in the Program Evaluation and Review Technique (PERT) have addressed a lack of statistical dependence modeling among activity duration uncertainties. However, their applications are hampered by two aspects: (1) the coherent monitoring of remaining project uncertainty as a project progresses by taking advantage of the degree of statistical dependence relies on complex computationally intensive procedures and (2) the specification of the degree of statistical dependence suffers from a curse of dimensionality in an application domain which already burdens experts with the estimation of activity most likely, lower and upper bound estimates. In this paper, we construct a continuous Bayesian Network (BN) model addressing both aspects by taking advantage of the BN inference procedure in the software AgenaRisk®. Specifically, the BN described defines a multivariate joint distribution between activity durations by incorporating only two additional dependence parameters to specify a degree of statistical dependence among the activities. Under certain dependence parameter settings, this BN model reduces to a multivariate joint distribution of statistically independent activities with the same marginal uncertainty description as the PERT method of Malcolm et. al (1959). To further facilitate application, an expert judgment elicitation procedure is developed to specify the two BN's dependence parameters via the elicitation of a sparse conditional median matrix of activity durations along a project network's paths. An illustrative example using a case study demonstrates the potential increased pace of learning about remaining project schedule uncertainty under a mild degree of statistical dependence by taking advantage of the Bayesian paradigm.

Keywords – **Project Schedule Risk, Bayesian Networks, Statistical Dependence, Uncertainty Modeling, Risk Analysis.**

---

*Corresponding author

Email addresses: ifenduka@gwmail.gwu.edu (Ifechukwu C. Nduka); dorpjr@gwu.edu (J. René van Dorp)

1 Introduction

Uncertainty is a characteristic feature of projects – this is empirically evident both in small and large projects. More so, project schedule uncertainty can be linked to impacts of external risk factors (for example, requirement change) which provide additional challenges for managing the project triple constraints of scope, schedule, and cost. The Standish Group (1995; 2009) notes that project risk factors and the resultant uncertainty they cause in project performance measures have contributed to project failures – costing loss of billions of dollars to project owners and sponsors. Today, with huge capital investments in projects, there is greater need for a project monitoring mechanism that will empower project managers with a robust toolkit for estimating project performance measures as a project progresses.

Recent advances in Program Evaluation and Review Technique (PERT) have addressed a lack of statistical dependence modeling among activity duration uncertainties in project schedule risk analysis (van Dorp J. R., 2005; van Dorp & Duffey, 1999; Khodakarami, Fenton, & Neil, 2007; Fang & Marle, 2012). Moreover, consideration of the questionable constant PERT variance assumption given an activity's range of support (Hahn, 2008) has recently led to the modified PERT variance procedure being suggested (Herrerias-Velasco et al., 2011) to further refine project schedule uncertainty assessment. The modified PERT variance is also influenced by an activity's most likely value, whereas in a classical PERT analysis an activity's variance is not.

The Bayesian Network (BN) approach towards modeling statistical dependence is the sole inference paradigm that allows for the ability to update/monitor project schedule uncertainty in a coherent manner while taking full advantage of the modeled statistical dependence, even for activities that are being completed but are not considered to be on a critical path. Specifically, in case of a statistical independence assumption among the activity durations, the completion of an activity in the project network that has a low criticality index (i.e. low probability of being on a critical path) will have very little or no effect on the remaining project completion time uncertainty. Conversely, if statistical activity-to-activity dependence is present, and modeled in the BN approach, the completion of that activity will affect duration uncertainty of activities that still need to be completed, by utilizing the BN inference procedure, and thus will indirectly influence the remaining project completion time uncertainty.

Such coherent monitoring of remaining project schedule uncertainty using the Bayesian paradigm as a project progresses through the completion of its activities has thus far, unfortunately, relied on a computationally intensive procedure called Markov Chain Monte Carlo (MCMC) sampling to obtain the updated project completion time distribution (Jenzarli, 1994); (Virto, Martin, & Insua, 2002); (Covaliu & Soyer, 1997); thereby limiting the practical implementation of these procedures. More recently, (Cho & Covaliu, 2003) and (Cho S. , 2009) designed a linear Bayes inference procedure to alleviate this computational complexity but had to sacrifice full distributional results and returned[3] to only providing updated mean and variance estimates of updated project completion time distributions. Moreover, the specification of the degree of statistical dependence among activities suffers from the curse of dimensionality even in moderately sized project networks. While some improvements have been suggested that address this aspect (Van Dorp, 2005), the elicitation burden to specify the degree of statistical dependence can still be considered substantial, further hampering the application of these methods.

In this paper, a continuous BN approach towards building statistical dependence among activities is presented, which allows one to obtain updated project completion time distribution results while addressing computational complexity by relying on the BN model's inference procedure implemented in AgenaRisk® (a specialized software tool for risk modeling and decision analysis with Bayesian Networks). Statistical dependence in discrete BNs is traditionally specified through the use of Conditional Probability Tables (CPTs) between probability nodes. The off-the-shelf software, AgenaRisk® , requires parametric inter-nodal relationships for the specification of statistical dependence in a continuous BN. While computationally convenient, this expediency of dependence specification through these parametric inter-nodal relationships has the disadvantage of a lesser transparency on how to specify the degree of dependence between nodes in the continuous BN than in a discrete BN using CPTs.

[3] In their classical paper, in which PERT originated, (Malcolm, Roseboom, Clark, & Fazar, 1959) also limited themselves to providing mean and variance estimates of project completion times.

The continuous BN developed herein utilizes novel inter-nodal parametric relationships in a manner that not only builds on the original PERT procedure by (Malcolm, Roseboom, Clark, & Fazar, 1959), but also details how a degree of statistical dependence among the activities changes by varying only two dependence parameters of the BN. In this process, we introduce a mode re-parameterization of the classical beta distribution while utilizing a two-sided power (TSP) distribution (Van Dorp & Kotz, 2002) as a prior distribution for the beta mode. A common beta distribution shape parameter and a common hyper TSP distribution power parameter, to be shared by all activities, will be demonstrated to drive both the degree of activity-to-activity statistical dependence and the marginal activity uncertainty. Both are important drivers for the prior completion time uncertainty in a project network, whereas the degree of inter-nodal dependence affects the speed of learning in a BN's posterior analysis procedure. The BN model requires the specification of the traditional PERT activity lower and upper bounds and a most likely estimate specification through expert judgement to describe activity completion uncertainty.

To further facilitate the specification of the degree of dependence between activities, an expert judgement elicitation procedure is proposed by eliciting activity-to-activity conditional medians which directly relates to a classical statistical dependence measure, Blomquist's q (Kruskal, 1958). (Garthwaite, Kadane, & O'Hagan, 2005) advocate the elicitation of probabilities over other statistical measures. The advantage of the proposed statistical elicitation procedure through probabilities is that it avoids eliciting correlation coefficients which are traditionally estimated from data, not using expert judgement. Finally, throughout this dependence elicitation procedure we introduce what we believe to be a novel distribution theory concept for capturing statistical dependence in a multivariate distribution termed *the conditional median matrix*. The elicitation burden for specifying the degree of activity-to-activity statistical dependence is reduced by limiting the elicitation of these conditional medians along a project network's paths resulting in a sparse conditional median matrix.

Subsequent sections of this paper are organized as follows: In Section 2 we discuss the parametric details of the continuous BN model introduced in this paper and the approach followed to implement the BN model in AgenaRisk®, using an illustrative example of a shipbuilding project (Taggart, 1980) consisting of 18 activities depicted in Figure 1 (at the duration level) and Figure 4. In Section 3, we explain how the degree of dependence between activities materializes in the BN through the specification of the common beta distribution shape parameter and a common hyper TSP distribution power parameter shared by all activities. Section 4 discusses degree of statistical dependence elicitation among the activities utilizing a conditional median approach and subsequent dependence parameter specification using a numerical procedure, which only has to be executed once. In Section 5, an uncertainty analysis is conducted with an illustrative case study using the BN framework depicted in Figure 1 and the example project network in Figure 4. Using what can be considered as a mild degree of statistical dependence between activities, and by coherently updating/monitoring the *a priori* project schedule uncertainty of the example network in Figure 4 with 95^{th} percentiles of durations as activities complete throughout the project network, we demonstrate the potential difference in speed of learning between utilizing the Bayesian paradigm and the more traditional approach of assuming statistical independence among the marginal distributions of activity durations. While, in general, incorporating statistical dependence *a priori* leads to larger uncertainty bands; *a posteriori* smaller uncertainty bands are observed after about a third into the project, in this case study. The faster learning about remaining completion time uncertainty combined with the precision of the BN approach may provide project managers more time to take corrective action to avoid schedule slippage. In Section 6, we summarize the insights gained from the analysis conducted in the illustrative example.

2 A Bayesian Network Dependence Model for Project Risk Analysis

A Bayesian Network activity dependence model in a project network enables coherent monitoring of project completion time uncertainty by utilizing the Bayesian paradigm. Figure 1 is a representation of the BN model developed herein which consists of three levels: The Common Quantile Level, the Mode Level, and the Duration Level. Statistical dependence among the marginal distributions of activity durations is modeled indirectly through the parent-child relationships between the uncertainty nodes represented in the BN in Figure 1.

i. *Common Quantile Level*: This level contains a single node Y that *a priori* represents the common quantile level for its child nodes in the Mode Level of the BN model in Figure 1. Hence, a priori $Y \sim U[0,1]$. *A posteriori*, this node provides for monitoring of overall project schedule performance. *A posteriori* levels for Y towards 1 (or 0) indicate typical activity completion above (or below) median activity durations.

ii. *Mode Level*: Given a project activity duration, X, with parameters: a_x = optimistic duration value; b_x = pessimistic duration value; and m_x = most likely value; the Mode Level provides an uncertainty model for the relative mode location Δ_x of each activity's mode given activity duration support $[a_x, b_x]$. Specifically, an activity's mode uncertainty is modeled *a priori* using a Two-Sided Power distribution, i.e. $\Delta_x \sim TSP(\delta_x, n)$ with probability density function (pdf) (van Dorp & Kotz, 2002):

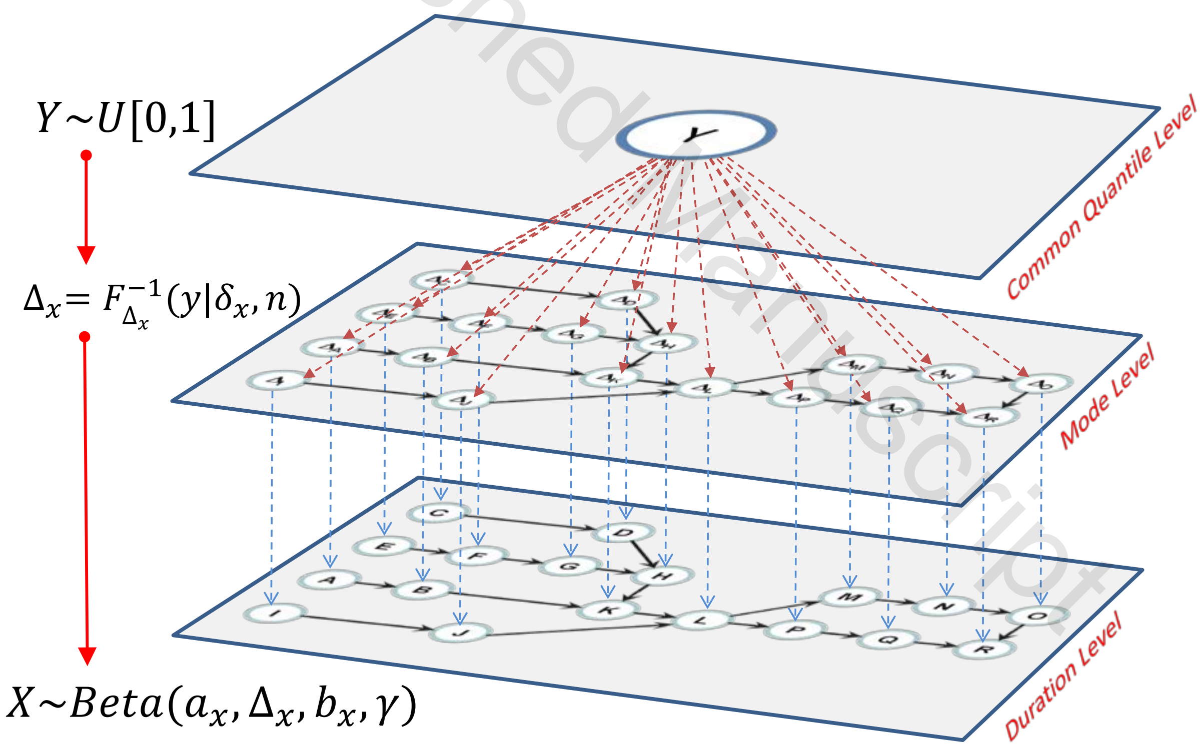


Figure 1: Bayesian Network (BN) model for project risk analysis. An example project network is depicted in the BN above at the duration level and the mode level.

$$f_{\Delta_x}(u|\delta_x, n) = \begin{cases} n\left(\dfrac{u}{\delta_x}\right)^{n-1}, & 0 \le u \le \delta_x, \\ n\left(\dfrac{1-u}{1-\delta_x}\right)^{n-1}, & \delta_x \le u \le 1, \end{cases} \quad (1)$$

where δ_x is the mode relative distance from the lower bound a_x given by

$$\delta_x = (m_x - a_x)/(b_x - a_x). \quad (2)$$

From (1), one obtains for the quantile function of the relative mode location Δ_x

$$F_{\Delta_x}^{-1}(y) = \begin{cases} \delta_x \sqrt[n]{y/\delta_x}, & 0 < y < \delta_x, \\ 1 - (1-\delta_x)\sqrt[n]{(1-y)/(1-\delta_x)}, & \delta_x < y < 1, \end{cases} \quad (3)$$

depicted in Figure 1.

iii. *Duration Level*: The prior activity uncertainty model for X is reminiscent of the classical PERT (Malcolm, 1959), except that herein $(X|a_x, b_x, \alpha_x, \beta) \sim Beta(a_x, b_x, \alpha_x, \beta)$ with pdf:

$$g_X(x|a_x, b_x, \alpha_x, \beta) = \frac{(b_x - a_x)^{-\beta}}{\mathcal{B}(\beta\alpha_x, \beta(1-\alpha_x))}(x-a_x)^{\beta\alpha_x - 1}(b_x - x)^{\beta(1-\alpha_x)-1}, \quad (4)$$

where $0 < \alpha_x < 1, \beta > 0$, $\mathcal{B}(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is the well-known beta normalization constant and

$$\alpha_x = \frac{\Delta_x(\beta-2)+1}{\beta}. \quad (5)$$

Substitution of (5) into (4) with $\gamma = \beta - 2$ leads to a re-parameterization of uni-modal beta distributions where the mode relative distance Δ_x is one of its parameters given by:

$$g_X(x|a_x, \Delta_x, b_x, \gamma) = \frac{(b_x - a_x)^{-(\gamma+1)}}{\mathcal{B}(\gamma\Delta_x + 1, \gamma(1-\Delta\alpha_x)+1)}(x-a_x)^{\gamma\Delta_x}(b_x - x)^{\gamma(1-\Delta_x)} \quad (6)$$

where $0 < \Delta_x < 1, \gamma > 0$. Thus the TSP distribution (1) serves as the prior distribution for the parameter Δ_x in equation (6). Throughout the remainder of this paper we shall use for simplicity the notation X to both refer to the activity itself and the random variable modeling its duration uncertainty.

Observe from (1) and (6) that the respective parameters n and γ are selected to be common among all activities. In Section 3, it shall be explained how the degree of statistical dependence between activities in the project network materializes through the specification of n and γ in (1) and (6) respectively.

2.1 The Bayesian Network Model Construction in AgenaRisk®

Construction of the Bayesian Network model in AgenaRisk® follows three major steps: First, we identify the set of variables or nodes that characterize the problem domain we want to investigate – thus we categorize nodes for the Common Quantile Level, the Mode Level, and the Duration Level. Second, we build the links across the nodes at the Duration Level to represent a project network. This is achieved by mapping sets of BN nodes that represent a project activity, and then using the forward pass of the Critical Path Method (CPM) of project scheduling to build the entire project network. Figure 2 shows a screenshot of the BN model in Figure 1 implemented in AgenaRisk®. Overall the BN model implementation of Figure 1 consists of 91 nodes and is thus a representation of a multivariate distribution of dimension 91.

The Common Quantile Level is represented in Figure 2 by the node $Y \sim U[0,1]$. The Mode Level consists of 36 nodes – for each activity X we have nodes D_x and LB_X (where $X = A, B, C, \cdots, R$) to implement the quantile function of the TSP distribution in (1) given by (3). The node LB_X implements a Boolean node to determine, based on the value of the random variable Y, if the lower or upper branch in (3) is selected. The node D_x evaluates $F^{-1}_{\Delta_x}(y)$ in (3) accordingly.

The Duration Level implements the CPM evaluation of a project network's completion time given activity duration values, and which in Figure 2 consists of 54 nodes. Each activity X is represented by a Start node, a Finish node and a Duration node as follows:

i. Start node: The *Start* node effectively represents the start of a project activity – modeled as a continuous node in AgenaRisk® using *Arithmetic Expression*, thus:

$$(Start)_x = Max[(Start)_y + (Duration)_y; for\ all\ preceding\ activities\ y] \qquad (7)$$

ii. Finish node: The *Finish* node represents completion time or end of a project activity – modeled as a continuous node in AgenaRisk® using *Arithmetic Expression*, thus:

$$(Finish)_x = (Start)_x + (Duration)_x \tag{8}$$

iii. Duration node: This represents the duration of a project activity – assumed to follow a Beta distribution with parameterization defined by Equation (6).

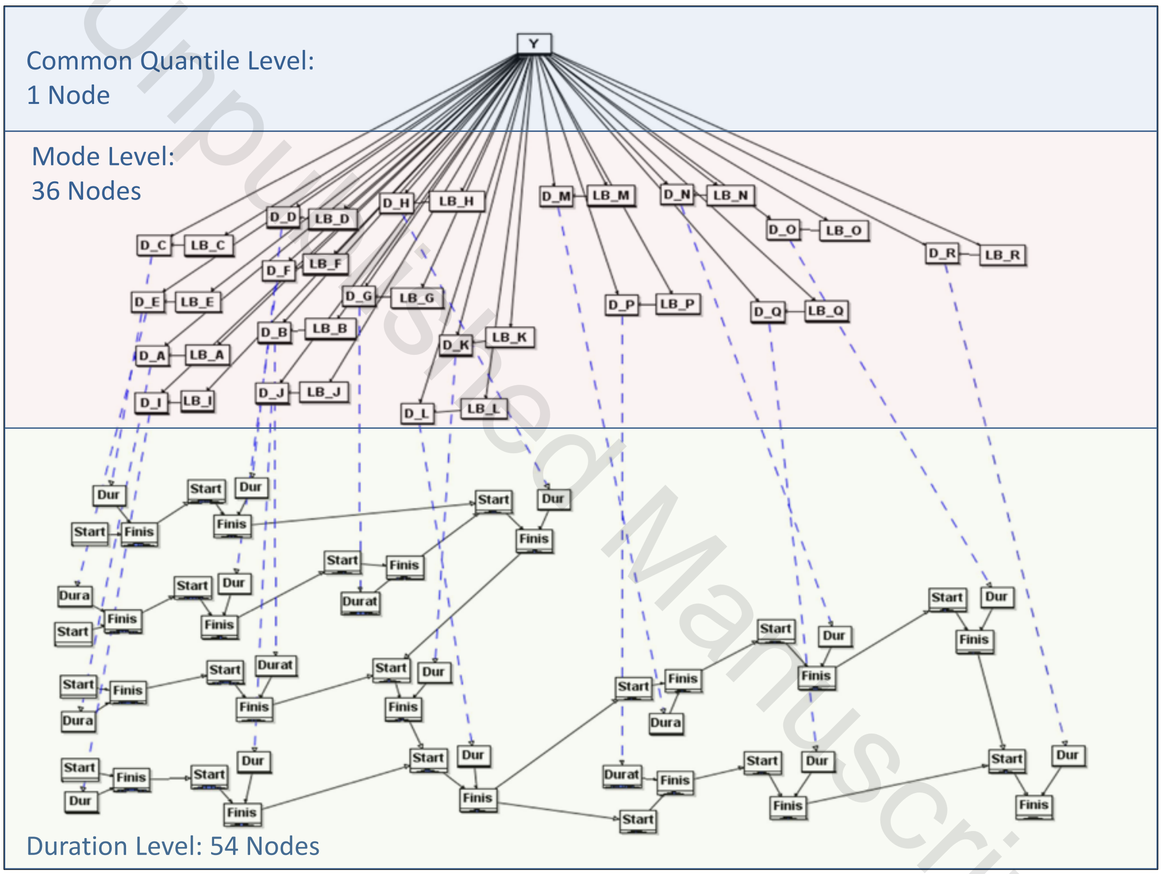


Figure 2: Bayesian Network model in Figure 1 implemented in AgenaRisk®

3 Varying the Degree of Statistical Dependence between Nodes of the BN in Figure 1.

Specifying the degree of dependence in a joint distribution in the absence of joint data necessitates a procedure geared towards its elicitation via expert judgment. In the case of a multivariate normal distribution the specification of a correlation matrix is necessary and in a discrete BN, CPTs are needed. Both, and other approaches towards constructing multivariate joint distributions, suffer from what is known to be “the curse of dimensionality” to specify the degree of dependence between its marginal random variables. For example, in case of the project network in Figure 1, a joint normal distribution approach towards specifying the degree of statistical dependence would require the specification of $\binom{18}{2} = 153$ additional correlation coefficients next to the activity duration data.

Efficient modeling of continuous BN nodes in AgenaRisk® (Neil, Tailor, & Marquez, 2007) allows for statistical dependence specification by defining parametric relationship between a parent node’s and child node’s distribution parameters, reducing the burden for dependence parameter specification considerably. In fact, the BN model in Figure 1 only requires the specification of 2 additional parameters n and γ in equations (1) and (6) that are common to all activities. Needless to say, this comes at a price in terms of flexibility of the statistical dependence structures that can be accommodated using this BN model. However, the BN model in Figure 1 was specifically designed with an emphasis towards reducing the elicitation burden of statistical dependence parameter specification among the activities in a PERT analysis, which currently hampers the application of statistical dependence among activities in a PERT context in practical settings. For example, a daunting task of specifying 153 correlation coefficients via expert judgement elicitation when using a joint normal distribution model for statistical dependence modeling in the 18 node activity network in Figure 1 may result in adopting, for the purpose of convenience only, the specious statistical independence assumption between activity durations.

Specification of n and γ in equations (1) and (6) determines the overall degree of statistical dependence specified amongst the activity durations in the BN in Figure 1, while also influencing the degree of uncertainty in the activity duration marginal distributions to a lesser extent. To explain this effect of the parameters n and γ in the multivariate joint distribution defined by the BN in Figure 1, we shall consider a series of illustrative examples of this BN using only two duration random variables for activities A and B, two mode random variables Δ_a and Δ_b and the common quantile random variable $Y \sim U[0,1]$. For example, using the parameter settings in Figure 3a, the Two-Sided Power distributions at the mode level converge to single point masses at modes δ_a and δ_b, blocking the influence of the node Y at the common quantile level while forcing the mode locations of the beta activity distributions A and B, at the duration level in Figure 3a, to be located at relative distances δ_a and δ_b of their support. The blocking at the mode level results in statistically independent beta marginal distributions with product moment correlation $\rho(A, B) = 0$ and mode relative distances for activities A and B at δ_a and δ_b, see also Equation (2). Hence, the BN parameter scenario for n and γ in Figure 3a reduces the BN model in Figure 1 to the original PERT analysis setting suggested by Malcolm et al. (1959).

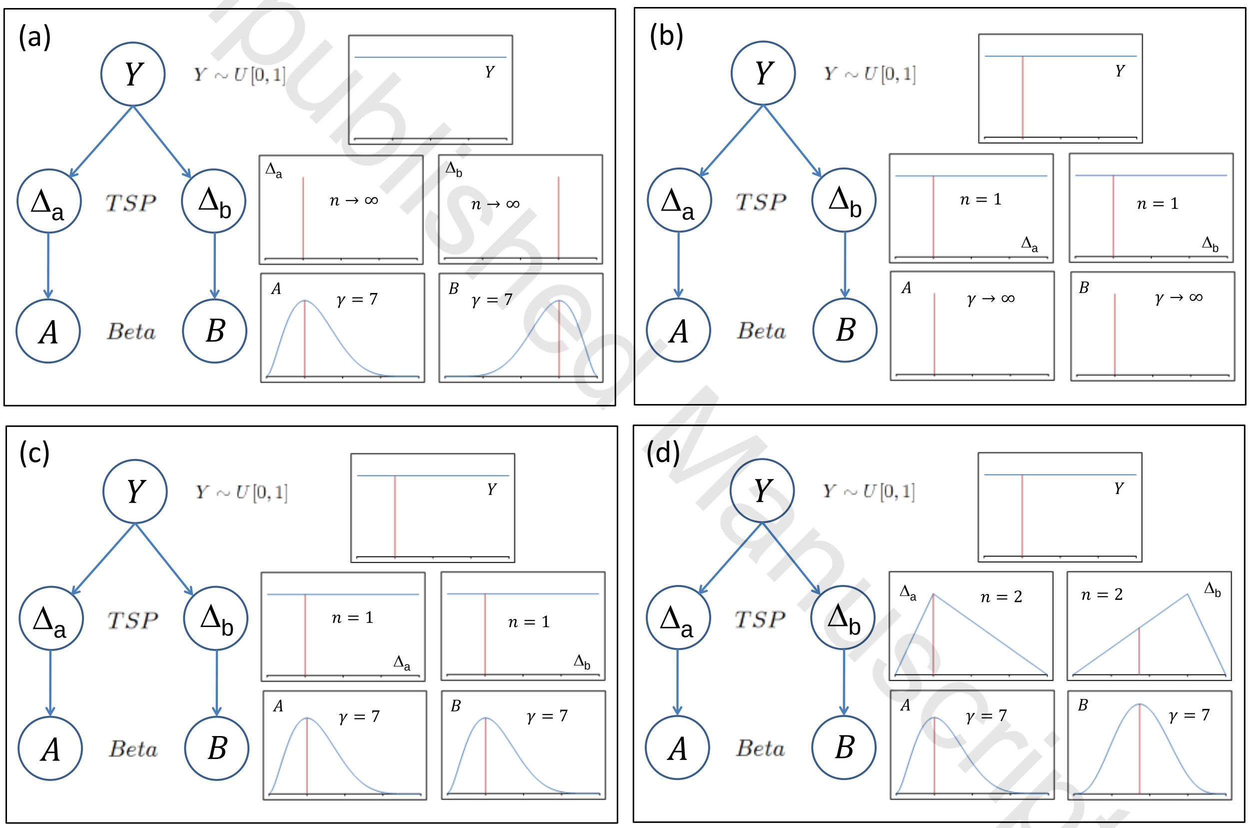


Figure 3: 5-node Bayesian Network (BN) examples of the BN Model in Figure 1;
(a) $\boldsymbol{\rho(A,B) = 0}$ (b) $\boldsymbol{\rho(A,B) = 1}$ (c) $\boldsymbol{0 < \rho(A,B) < 1}$, (d) $\boldsymbol{0 < \rho(A,B) < 1}$.

Another extreme is depicted in Figure 3b. In contrast to Figure 3a, the beta distributions at the duration level converge to single point masses at realizations of the mode random variables Δ_a and Δ_b (indicated by the vertical lines at the mode level in Figure 3b). By setting $n = 1$ in Figure

3b at the mode level, the prior TSP distributions reduce to uniform distributions on $[0,1]$ whose realizations will be equal to the realization $y = 0.25$ depicted in Figure 3b of the random variable $Y \sim U[0,1]$ at the common quantile level. Thus, in Figure 3b the prior marginal distributions of the activity durations A and B reduce to uniform distributions with product moment correlation $\rho(A,B) = 1$. Hence, the BN scenario depicted in Figure 3b reduces the BN model in Figure 1 to a PERT analysis with uniformly distributed activity durations with pairwise correlations equal to 1 and thus *a priori* results in a project completion time distribution with the largest possible uncertainty given the activity's most likely and lower and upper bound estimates. On the other hand, because of the pairwise correlations equal 1, the scenario depicted in Figure 3b also coincides with maximum *a posteriori* learning.

An intermediate scenario between Figure 3a and Figure 3b is depicted in Figure 3c. Here the TSP distributions at the mode level are still uniform, but the random variables at the duration level are beta distributed. Note that given the realization at the common quantile level and the uniformity at the mode level, the beta distribution for activity B has switched at the duration level to being the same as that of activity A (please compare Figure 3c with Figure 3a to notice that switching). As a result of the setup in Figure 3c, the mode locations of the activity durations A and B will now "move" in sync resulting in a positive correlation $\rho(A,B) > 0$ in Figure 3c. The larger the value of γ in Figure 3c, the more "peaked" the beta distribution will be at the duration level of the BN, but will still share the same mode, resulting in a larger correlation $\rho(A,B)$ as γ increases. In the limit, if we let $\gamma \to \infty$ the BN scenario in Figure 3c converges to the situation in Figure 3b with $\rho(A,B) \to 1$.

Finally, Figure 3d depicts a similar scenario as in Figure 3c, but where $n = 2$ changes the uniform distribution at the mode level in Figure 3c to a triangular distribution in Figure 3d. Please note that the beta distribution for activity B has a mode that has now shifted to the right as compared to the mode location in Figure 3c for activity B. Hence, as a result of the larger value of n in Figure 3d, the mode locations of the activities A and B are less affected by mode level realizations in the BN, resulting in a lower correlation $\rho(A,B)$ in Figure 3d as compared to

Figure 3c. In the limit, if we let $n \to \infty$ the BN scenario in Figure 3d converges to the situation in Figure 3a with $\rho(A, B) \to 0$.

Summarizing, by increasing the beta parameter γ in (6) while keeping the TSP power parameter n in (1) constant, the degree of statistical dependence among the activities in the BN in Figure 1 increases. Conversely, by decreasing the TSP power parameter n in (1) while keeping the beta parameter γ in (1) constant, the degree of statistical dependence among the activities in the BN in Figure 1 decreases.

In general, in a BN approach where parent-child dependence is specified parametrically for continuous nodes, or through the use of Conditional Probability Tables (CPTs) for discrete nodes, the marginal distributions of nodes in the BN follow as an output instead of the marginal distributions being specified as an input (as is the case in a classical PERT analysis). Hence part of the challenge in specifying values for the parameters n and γ is to select values such that the variance of the activity durations are close to the modified PERT variance values in Table 1 (or the PERT variance values, should one prefer the classical PERT approach) while satisfying a degree of statistical dependence requirement. In Section 4, we develop such a procedure for the BN in Figure 1.

4 Degree of Statistical Dependence Elicitation

Pearson's product-moment correlation, Spearman's rank correlation or Kendall's tau while suited for measuring the degree of dependence between two random variables, do not lend themselves well for expert judgment elicitation. Despite (Kruskal, 1958) having provided a decision analytic explanation of these statistical dependence measures, their explanations demonstrate the level of cognitive processing required to elicit them either directly or indirectly. However, (Kruskal, 1958) also discusses the dependence measure Blomquist's q proposed by (Blomquist, 1950), which is less well known, but is directly related to a conditional median of one of the random variables given the other's median value. Elicitation of probabilities has been proposed and studied extensively (see Garthwaite et al. (2005) for an excellent overview). Therefore, we suggest the elicitation of conditional medians between activity durations in the project network as an indirect elicitation procedure towards specification of their common parameters n and γ in equations (1) and (6), respectively.

An example question in the elicitation procedure is: *Suppose, for example, Activity A has finished above its median value $a_{0.5}$, what is the probability that Activity B finishes above its median value $b_{0.5}$?* If an expert's answer is:

- "= 0.50," Activity A and Activity B are modeled as statistically independent
- in (0.50, 1.0), Activity A and Activity B are modeled as positively dependent[4]
- " = 1.00," Activity A determines Activity B

In further developing this procedure, we refer to Figure 4, which is the same as our example project in Figure 1 at the duration level with letters representing the project activities (see also Table 1).

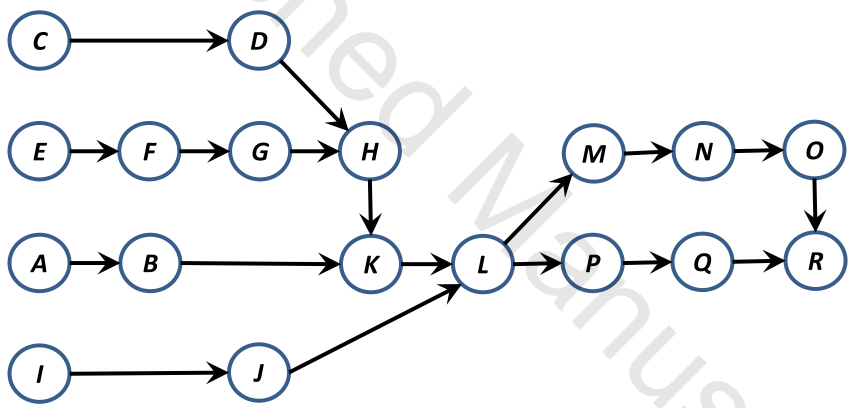


Figure 4: Example project (represented as an Activity-on-Node project network)

To further specify the degree of dependence in the BN model in Figure 1, the elicitation of Activity-to-Activity conditional medians is suggested for the project in Figure 4 that are only path-dependent as defined by the project network. Thus, for example, we propose elicitation of:

[4] Two random variables are positively dependent when high (small) values of one tend to be associated with high (small) values of the other.

$$\Pr\{B > b_{0.5} | A > a_{0.5}\}, \Pr\{K > k_{0.5} | B > b_{0.5}\}, \Pr\{L > l_{0.5} | K > k_{0.5}\}, \quad (9)$$
$$\Pr\{P > p_{0.5} | L > l_{0.5}\}, \Pr\{Q > q_{0.5} | P > p_{0.5}\}, \Pr\{R > r_{0.5} | Q > q_{0.5}\}$$

which completes the elicitation of Activity-to-Activity conditional medians along one of the eight paths in Figure 4. A reader can easily verify that the completion of this path-dependent Activity-to-Activity conditional median elicitation procedure leads to an elicitation requirement of only 18 Activity-to-Activity conditional medians. The 18 Activity-to-Activity conditional medians to be used in our illustrative example are provided in Figure 5 in a conditional median matrix.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	1.00	$(B\|A)$																
B	0.65	1.00									$(K\|B)$							
C			1.00	$(D\|C)$														
D			0.65	1.00				$(H\|D)$										
E					1.00	$(F\|E)$												
F					0.70	1.00	$(G\|F)$											
G						0.65	1.00	$(H\|G)$										
H				0.70			0.60	1.00			$(K\|H)$							
I									1.00	$(J\|I)$								
J									0.70	1.00		$(L\|J)$						
K		0.75						0.65			1.00	$(L\|K)$						
L										0.65	0.60	1.00	$(M\|L)$			$(P\|K)$		
M												0.70	1.00	$(N\|M)$				
N													0.55	1.00	$(O\|N)$			
O														0.70	1.00			$(R\|O)$
P												0.65				1.00	$(Q\|P)$	
Q																0.65	1.00	$(R\|Q)$
R															0.70		0.60	1.00

Figure 5: Conditional median matrix showing path-dependent Activity-to-Activity conditional medians for the project in Figure 4.

Observe from:

$$\Pr(Y > y_{0.5} | X > x_{0.5}) = \frac{\Pr(\mathrm{X} > x_{0.5} | \mathrm{Y} > y_{0.5}) \Pr(Y > y_{0.5})}{\Pr(X > x_{0.5})} = \Pr(\mathrm{X} > x_{0.5} | \mathrm{Y} > y_{0.5}) \quad (10)$$

that the conditional median matrix in Figure 5 is symmetric across the diagonal by definition. The upper (lower) half above the diagonal lists the 18 Activity-to-Activity conditional median definitions (values).

4.1 Activity Uncertainty Specification

Activity specific information for the example project network in Figure 1 and Figure 4 is provided in Table 1. Table 1 evaluates by activity X the classical PERT Variance

$$(b_x - a_x)^2/36 \tag{11}$$

and the modified PERT variance

$$C(\delta_x)(b_x - a_x)^2/36, \tag{12}$$

where

$$C(\delta_x) = [5 + 16\delta_x(1 - \delta_x)]/7. \tag{13}$$

Table 1: Activity duration data for the project network depicted in Figure 1 and Figure 4

Activity Description	ID	a	m	b	δ	C(δ)	PERT Variance	Modified PERT Variance
Layout Bottom Shell	A	22	25	30	0.375	1.250	1.778	2.222
Assemble Bottom Shell	B	35	37	43	0.250	1.143	1.778	2.032
IB Piping Layout	C	19	22	29	0.300	1.194	2.778	3.317
IB Piping Fabricate	D	4	5	10	0.167	1.032	1.000	1.032
IB Struc. Layout	E	23	26	31	0.375	1.250	1.778	2.222
IB Struct. Fabricate	F	16	18	24	0.250	1.143	1.778	2.032
IB Struct. Assemble	G	11	14	20	0.333	1.222	2.250	2.750
IB Piping Install	H	6	7	12	0.167	1.032	1.000	1.032
Main Engine Found. Layout	I	25	28	33	0.375	1.250	1.778	2.222
Main Engine Found. Fabricate	J	33	35	40	0.286	1.181	1.361	1.607
Erect IB	K	27	30	37	0.300	1.194	2.778	3.317
Erect Foundation	L	6	7	11	0.200	1.080	0.694	0.750
Complete 3rd Deck	M	4	5	9	0.200	1.080	0.694	0.750
Install Boiler	N	6	7	10	0.250	1.143	0.444	0.508
Test Boiler	O	9	10	15	0.167	1.032	1.000	1.032
Install Main Engine	P	6	7	12	0.167	1.032	1.000	1.032
Finish Engine	Q	17	20	26	0.333	1.222	2.250	2.750
Final Test	R	13	15	20	0.286	1.181	1.361	1.607
					Average Variance		1.528	1.790

In (12) and (13), δ_x is given by (2) and $C(\delta_x)$ is interpreted as the PERT variance adjustment factor accounting for the relative location δ_x of m_x in the range $[a_x, b_x]$ of Activity X, where $X = A, B, C, \cdots, R$. Observe from Table 1 that for all activities the PERT variance adjustment factor $C(\delta_x)$ is larger than one, implying a larger duration uncertainty for these parameter settings as compared to using the classical PERT variance (11).

4.2 Solving for the Degree of Dependence Parameters n and γ in Equations (1) and (6)

From the modified PERT variance column in Table 1, one evaluates an average variance of 1.79 and from the conditional medians in Figure 5 an average conditional median of 0.66. Below we shall describe a numerical procedure to solve for the parameter n in (1) and the parameter γ in (6) such that prior average variance across activities *approximately* equates to the value 1.79 and the average of activity-to-activity conditional medians *approximately* equates to the value 0.66. Figure 6a displays the behavior of average activity variance as a function of γ and n, whereas Figure 6b displays the behavior of average activity-to-activity conditional median. Observe from Figure 6a that average activity variance decreases with increasing values of γ and n. On the other hand, one observes from Figure 6b an increasing (decreasing) behavior in the average activity-to-activity conditional median as the parameter γ (parameter n) increases.

Figure 6 was generated by sampling from the prior distribution of the BN model in Figure 1 for different values of γ and n, separate from the BN model's implementation in AgenaRisk®. Figure 7 displays the data for Figure 6 containing 441 cell values for Figure 6a and Figure 6b. Each cell value in Figure 7 was evaluated by generating 1000 joint prior samples from the BN Model in Figure 1. Thus the generation of Figure 7 involved 441,000 joint samples of dimension 18 totaling about 8 million samples from the activity duration marginal distributions. The top part of Figure 7 displays in the shaded region values of the average activity variance within the range $(1.50, 2.00)$. The bottom part of Figure 7 displays in the shaded region values of the average activity-to-activity conditional medians within the range $(0.625, 0.675)$. By intersecting the shaded regions, reminiscent of intersecting iso-contours, we select the values $n = 3.5$ and $\gamma = 11$ by identifying cells with an average activity variance of 1.69 and an average activity-to-activity conditional median of 0.66.

Next, evaluating the values of average activity variance and average activity-to-activity conditional median by sampling once more one thousand times from the prior distribution of the BN model in Figure 1, we evaluate remaining epistemic 90% probability interval (1.69, 1.83) of average activity variance and remaining epistemic 90% probability interval (0.652, 0.671) for average activity-to-activity conditional median. The first and second interval [(1.69, 1.83) and

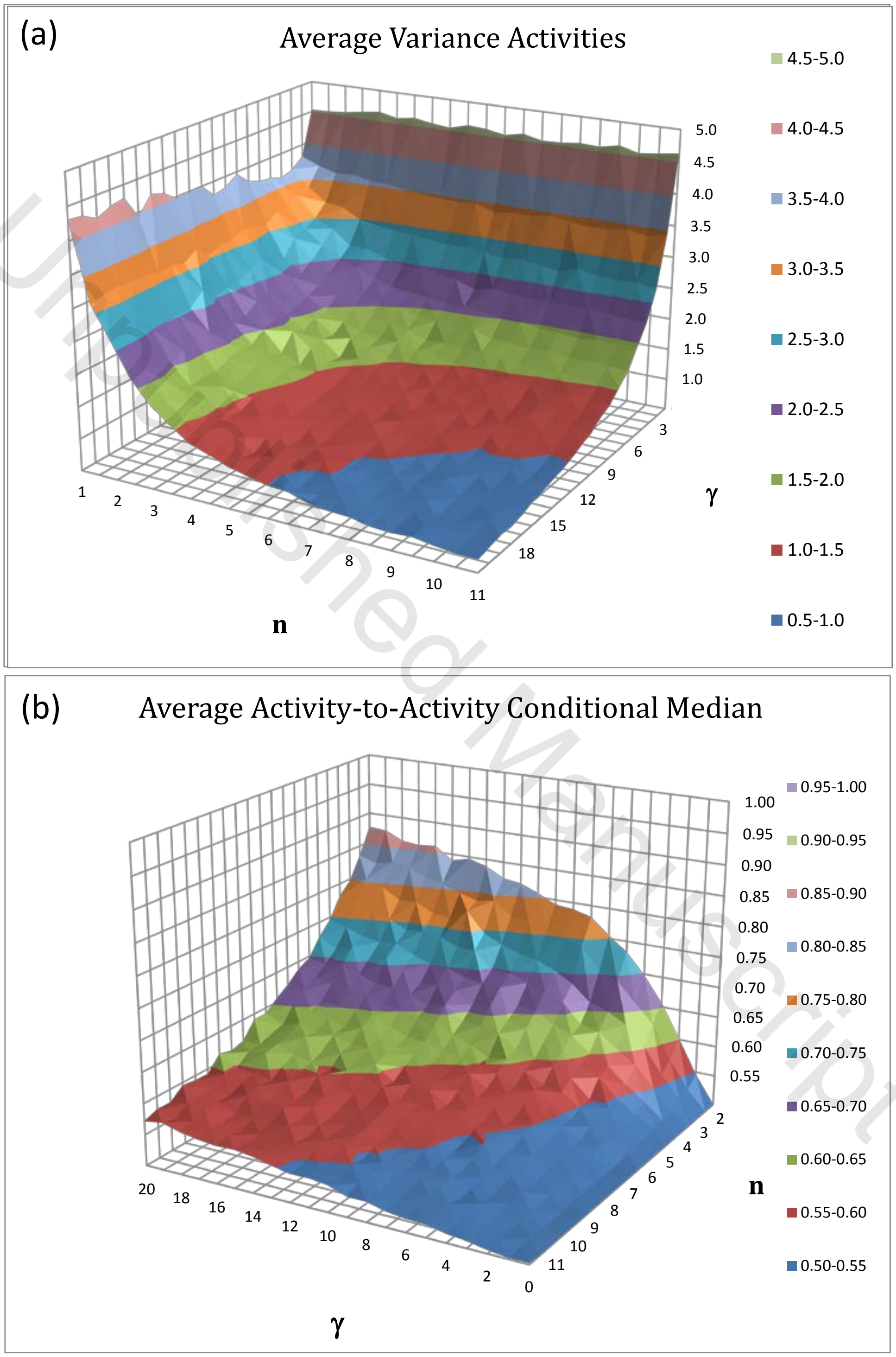


Figure 6: (a) Average Prior Marginal Activity Variance as a function $\boldsymbol{\gamma}$ and $\boldsymbol{n}$ and
(b) Average Prior Activity-to-Activity Conditional Median as a function $\boldsymbol{\gamma}$ and $\boldsymbol{n}$

γ →

n ↓

Prior Average Activity Variance between 1.50 and 2.00

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4.53	3.80	3.65	3.59	3.66	3.69	3.77	3.96	3.83	3.78	4.00	3.98	3.97	4.13	4.19	3.93	4.30	4.26	4.16	4.28	4.31
2	4.54	3.69	3.40	3.34	3.23	3.23	3.10	3.17	3.19	3.12	3.10	3.04	3.22	3.09	3.14	3.12	3.18	3.17	3.20	3.17	3.11
2	4.56	3.61	3.19	3.01	2.90	2.83	2.59	2.61	2.75	2.67	2.56	2.57	2.52	2.72	2.53	2.53	2.49	2.50	2.61	2.61	2.62
3	4.61	3.62	3.15	2.87	2.68	2.52	2.40	2.42	2.35	2.33	2.30	2.17	2.16	2.13	2.22	2.13	2.13	2.03	2.11	2.15	2.12
3	4.65	3.57	2.99	2.64	2.49	2.37	2.25	2.15	2.09	2.04	2.06	2.00	1.87	1.89	1.83	1.75	1.78	1.74	1.84	1.79	1.78
4	4.58	3.56	2.94	2.64	2.42	2.25	2.12	1.96	1.95	1.85	1.93	1.69	1.74	1.73	1.65	1.60	1.69	1.58	1.59	1.52	1.56
4	4.62	3.51	2.91	2.54	2.28	2.13	2.01	1.84	1.73	1.75	1.71	1.55	1.61	1.57	1.51	1.46	1.38	1.40	1.40	1.35	1.35
5	4.57	3.48	2.88	2.46	2.19	2.07	1.89	1.75	1.73	1.53	1.56	1.49	1.48	1.45	1.39	1.35	1.35	1.31	1.27	1.23	1.25
5	4.54	3.46	2.86	2.49	2.15	2.00	1.84	1.68	1.64	1.56	1.44	1.36	1.35	1.30	1.27	1.28	1.22	1.21	1.10	1.14	1.14
6	4.62	3.44	2.82	2.43	2.13	1.91	1.78	1.69	1.53	1.47	1.37	1.33	1.31	1.24	1.21	1.18	1.11	1.12	1.11	1.07	1.06
6	4.60	3.46	2.83	2.39	2.08	1.90	1.67	1.61	1.48	1.41	1.34	1.27	1.23	1.21	1.14	1.10	1.09	1.05	1.02	0.99	1.00
7	4.55	3.37	2.80	2.27	2.07	1.88	1.70	1.56	1.46	1.35	1.31	1.23	1.16	1.18	1.11	1.08	1.04	1.07	0.96	0.91	0.98
7	4.61	3.41	2.75	2.35	2.05	1.87	1.67	1.52	1.40	1.35	1.23	1.20	1.17	1.09	1.09	1.01	0.99	0.99	0.97	0.89	0.88
8	4.54	3.41	2.78	2.31	2.01	1.77	1.67	1.48	1.40	1.30	1.22	1.19	1.12	1.04	1.03	1.00	0.96	0.91	0.92	0.89	0.85
8	4.55	3.46	2.74	2.30	2.01	1.77	1.62	1.45	1.39	1.26	1.20	1.12	1.08	1.01	1.00	0.93	0.94	0.87	0.85	0.84	0.83
9	4.58	3.41	2.69	2.30	1.95	1.73	1.59	1.45	1.35	1.24	1.14	1.13	1.05	1.01	0.94	0.91	0.89	0.86	0.82	0.80	0.77
9	4.57	3.38	2.79	2.30	1.95	1.74	1.57	1.42	1.28	1.21	1.16	1.10	1.00	0.97	0.94	0.93	0.87	0.85	0.80	0.77	0.76
10	4.63	3.39	2.73	2.29	1.91	1.70	1.55	1.40	1.31	1.24	1.14	1.06	1.02	0.96	0.93	0.85	0.83	0.78	0.78	0.80	0.78
10	4.55	3.43	2.71	2.27	1.95	1.71	1.53	1.42	1.29	1.19	1.13	1.03	1.01	0.95	0.89	0.85	0.82	0.77	0.77	0.76	0.73
11	4.59	3.37	2.73	2.24	1.95	1.70	1.51	1.36	1.26	1.17	1.09	1.01	0.94	0.91	0.89	0.85	0.81	0.79	0.76	0.72	0.69
11	4.63	3.45	2.67	2.24	1.96	1.68	1.53	1.36	1.28	1.16	1.10	1.02	0.96	0.93	0.88	0.85	0.79	0.75	0.76	0.73	0.69

Prior Average Conditional Median between 0.625 and 0.675

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.50	0.55	0.60	0.66	0.70	0.73	0.75	0.78	0.79	0.79	0.81	0.82	0.82	0.84	0.85	0.84	0.86	0.86	0.86	0.87	0.88
2	0.50	0.53	0.58	0.63	0.66	0.69	0.71	0.74	0.75	0.75	0.77	0.77	0.79	0.79	0.80	0.81	0.81	0.83	0.83	0.82	0.83
2	0.50	0.53	0.56	0.60	0.63	0.65	0.66	0.69	0.70	0.72	0.72	0.73	0.75	0.78	0.76	0.77	0.78	0.78	0.80	0.80	0.80
3	0.50	0.52	0.55	0.58	0.60	0.63	0.63	0.66	0.68	0.69	0.70	0.70	0.72	0.73	0.74	0.74	0.74	0.75	0.76	0.77	0.78
3	0.50	0.51	0.54	0.56	0.59	0.60	0.62	0.64	0.65	0.66	0.68	0.68	0.69	0.70	0.70	0.71	0.73	0.72	0.73	0.74	0.75
4	0.50	0.51	0.53	0.55	0.58	0.59	0.61	0.62	0.63	0.64	0.66	0.66	0.68	0.68	0.68	0.69	0.71	0.70	0.72	0.71	0.73
4	0.50	0.51	0.52	0.54	0.55	0.58	0.59	0.60	0.61	0.63	0.64	0.64	0.66	0.66	0.67	0.67	0.68	0.69	0.70	0.69	0.70
5	0.50	0.51	0.52	0.54	0.55	0.57	0.57	0.59	0.61	0.60	0.62	0.63	0.64	0.64	0.65	0.66	0.66	0.67	0.67	0.68	0.69
5	0.50	0.51	0.52	0.54	0.54	0.56	0.58	0.58	0.59	0.60	0.60	0.61	0.62	0.62	0.63	0.65	0.65	0.65	0.65	0.67	0.67
6	0.50	0.50	0.52	0.53	0.54	0.55	0.56	0.58	0.58	0.60	0.60	0.60	0.61	0.62	0.63	0.64	0.63	0.64	0.65	0.65	0.66
6	0.50	0.50	0.52	0.53	0.54	0.55	0.55	0.56	0.57	0.58	0.59	0.59	0.60	0.61	0.62	0.62	0.63	0.63	0.64	0.63	0.64
7	0.50	0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.57	0.58	0.59	0.59	0.61	0.61	0.61	0.61	0.63	0.63	0.62	0.64
7	0.50	0.50	0.52	0.52	0.53	0.53	0.55	0.55	0.56	0.57	0.57	0.58	0.59	0.59	0.60	0.59	0.60	0.61	0.62	0.62	0.61
8	0.50	0.50	0.51	0.52	0.53	0.54	0.55	0.55	0.55	0.56	0.57	0.58	0.58	0.58	0.59	0.60	0.60	0.60	0.62	0.61	0.61
8	0.50	0.50	0.51	0.52	0.52	0.53	0.54	0.54	0.55	0.56	0.56	0.57	0.57	0.57	0.58	0.58	0.59	0.59	0.59	0.60	0.61
9	0.50	0.50	0.51	0.52	0.52	0.53	0.53	0.54	0.55	0.55	0.56	0.57	0.56	0.57	0.58	0.58	0.58	0.59	0.59	0.60	0.59
9	0.50	0.50	0.51	0.51	0.52	0.52	0.53	0.53	0.54	0.54	0.55	0.56	0.56	0.57	0.57	0.58	0.58	0.59	0.58	0.59	0.59
10	0.50	0.50	0.51	0.51	0.52	0.52	0.53	0.54	0.54	0.55	0.55	0.56	0.56	0.56	0.57	0.57	0.57	0.57	0.58	0.58	0.60
10	0.50	0.50	0.51	0.51	0.52	0.53	0.53	0.53	0.53	0.54	0.55	0.55	0.56	0.56	0.56	0.56	0.56	0.57	0.58	0.58	0.58
11	0.50	0.50	0.51	0.51	0.52	0.52	0.53	0.53	0.53	0.54	0.54	0.54	0.55	0.55	0.56	0.56	0.57	0.57	0.57	0.57	0.58
11	0.50	0.50	0.51	0.51	0.51	0.52	0.52	0.52	0.53	0.53	0.54	0.55	0.55	0.55	0.56	0.56	0.56	0.56	0.57	0.58	0.57

Figure 7: Data for the 3D graphs in Figure 6

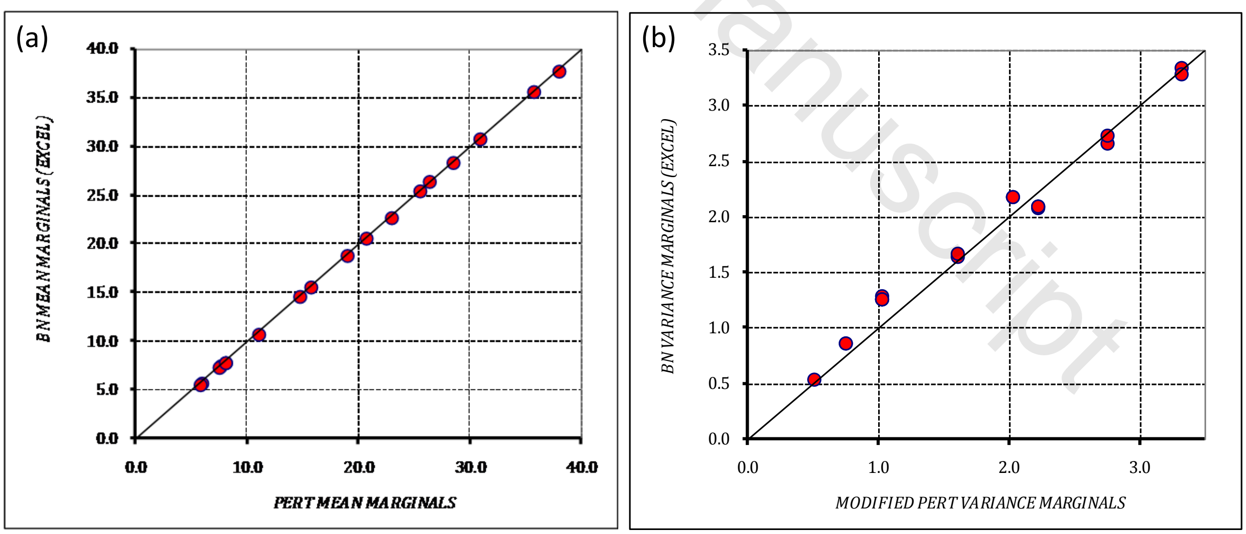


Figure 8: Visual comparison of (a) BN activity mean and PERT Mean and (b) BN activity variance and PERT modified activity variance for the eighteen activities listed in Table 1 using parameter values $\boldsymbol{n} = \mathbf{3.5}$ and $\boldsymbol{\gamma} = \mathbf{11}$

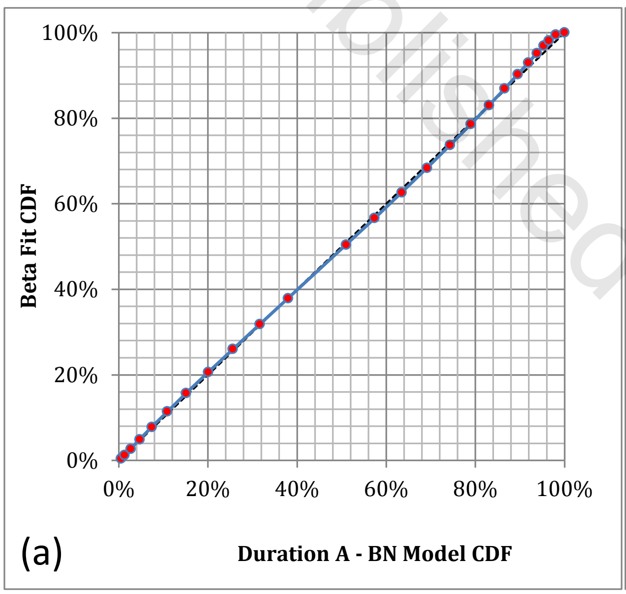
(0.652, 0.671)] contain the value 1.79 and 0.66, respectively. The value 1.79 was evaluated from the modified PERT variance column in Table 1, and 0.66 was evaluated from the conditional median matrix in Figure 5. In terms of Pearson's product-moment correlation, we evaluate the remaining epistemic 90% probability interval (0.487, 0.530) of the average activity-to-activity correlation, which indicates a mild degree of activity-to-activity positive dependence.

Finally for verification, using the same prior sampling procedure, we evaluate activity means and variances of the BN model in Figure 1 using the values $n = 3.5$ and $\gamma = 11$ and the data in Table 1. Figure 8a above compares visually the activity duration BN means and PERT means from Table 1 while Figure 8b compares visually activity duration BN variances with the modified activity PERT variances provided in Table 1. Both Figure 8a and Figure 8b demonstrate a close match for both the mean values and variances across all eighteen activity durations.

5 Monitoring Project Uncertainty as the Project Progresses

The purpose of this illustrative example is to demonstrate the potential benefit of modeling statistical dependence among activity durations on the pace of learning about remaining completion time uncertainty when a project progresses, utilizing the Bayesian paradigm. To that end, we shall compare remaining project completion uncertainty at various milestones (or activity completions) using the BN model in Figure 1 (in which statistical dependence is incorporated) with an approach where statistical independence among the marginal distributions of activity durations is assumed. It is important to note that the approach with an assumption of statistical independence among the marginal distributions of activity durations, is also implemented based on the example project of Figure 1. The analysis and implementation of this statistical independence case in AgenaRisk®, however, includes only the duration level elements of the BN model in Figure 1.

To allow for the aforementioned comparison and to ensure we are evaluating the effect of statistical dependence in the BN Model in Figure 1 on project schedule uncertainty, and not the effect of a difference in marginal distribution uncertainties, the prior marginal activity duration distributions were generated from the BN model (with statistical dependence) in Figure 1 using the same sampling procedure described in Section 3. Next, beta distributions were fitted to the sampled prior activity distributions using the least squares method to serve as the marginal activity duration distributions for the statistical independence case. Figure 9 contains two probability-probability (P-P) plots comparing visually the prior sampled cdf values of Activities A and B with those evaluated using their fitted beta distributions. In case of a perfect fit, the P-P plots would fall on the unit diagonal. Observe from Figure 9 that the beta fitted distributions match well with prior sampled distributions of Activity A and Activity B. P-P plots for the other 16 activity durations that appear as nodes in Figure 1 show similar behavior. Table 2 lists the parameters of the beta fitted marginal distributions of the activity durations in the BN model in Figure 1. In addition, the last row in Table 2 provides the 95th percentiles of these BN marginal distributions.



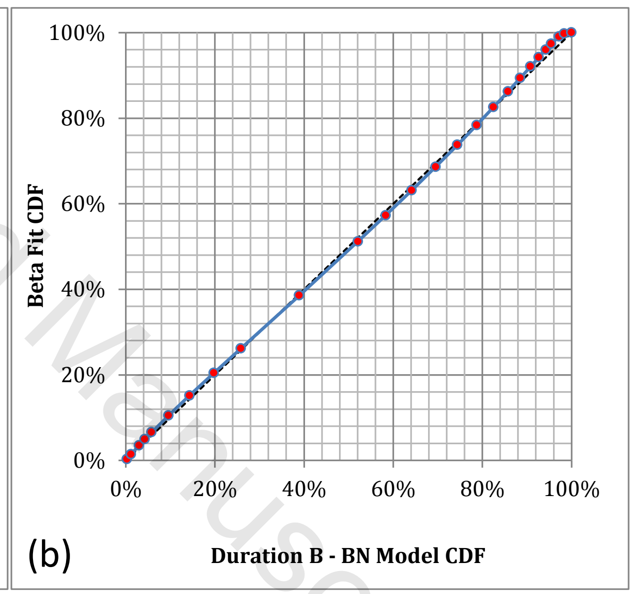


Figure 9: P-P plots comparing BN marginal distributions with Least Squares Beta fitted distributions for (a) Duration Activity A (b) Duration Activity B

Table 2: Parameters of Beta fitted marginal distributions of activity durations and their 95th percentiles

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
LB	22	35	19	4	23	16	11	6	25	33	27	6	4	6	9	6	17	13
UB	30	43	29	10	31	24	20	12	33	40	37	11	9	10	15	12	26	20
alpha	2.95	2.41	2.63	2.01	2.95	2.41	2.77	2.03	2.96	2.58	2.65	2.18	2.19	2.41	2.02	2.03	2.77	2.58
beta	3.73	3.89	3.85	3.86	3.73	3.89	3.81	3.88	3.74	3.88	3.87	3.90	3.91	3.89	3.88	3.88	3.81	3.88
95th	27.97	40.60	26.18	8.02	28.97	21.60	17.58	10.02	30.97	37.99	34.18	9.41	7.41	8.80	13.02	10.02	23.58	17.99

In this example, we shall instantiate the BN models with these 95th percentiles representing activity completions for both the statistical dependence and statistical independence cases. Subsequently, the state-of-the-art Bayesian inference algorithm implemented in AgenaRisk® enables propagation of influence across the nodes of the BN models, and also evaluates the remaining completion time distributions as the project progresses. Figure 10a and Figure 10b depict the CPM evaluation of the example project in Figure 1 using the activity duration most likely values provided in Table 1 and the activity duration 95th percentiles provided in Table 2. A CPM evaluation using only activity most likely values results in a project completion time of 144 days (depicted in Figure 10a). When one uses only activity 95th percentile values, a project completion time of 173.4 days (depicted in Figure 10b) is evaluated using the CPM. In addition, four intermediate milestones are identified in Figure 10a and Figure 10b.

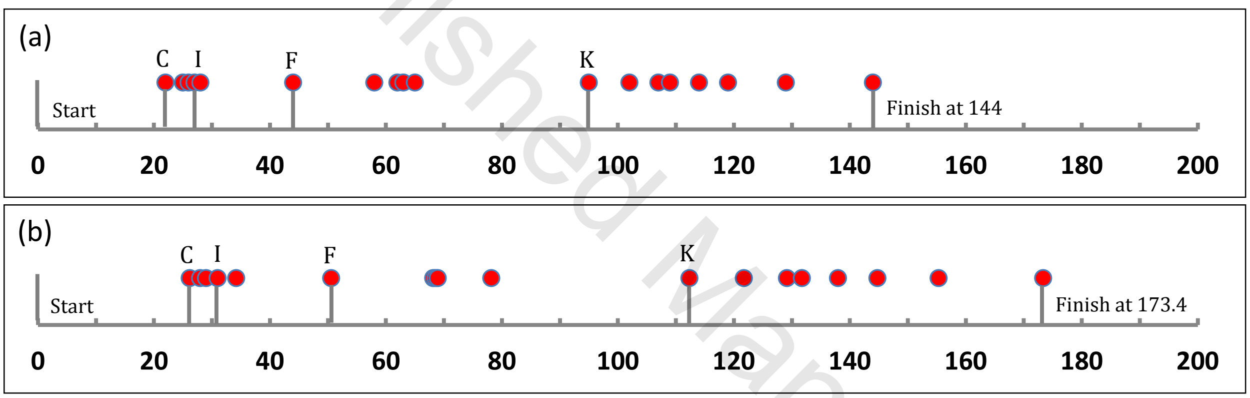


Figure 10: CPM evaluation of the example project in Figure 1 using - (a) activity duration most likely values shown in Table 1 (b) using 95th percentiles of activity durations shown in Table 2.

The first milestone captures the completion of the first activity (Activity C), the second milestone coincides with the completion of Activity I and captures the completion of all activities that do not have a predecessor activity (see Figure 4). The third and fourth intermediate milestones coincide with the completion of activities at about one-third (Activity F) and two-thirds (Activity K) into the project. We shall evaluate the project completion time uncertainty distribution at the start of the project as well as the remaining project completion time uncertainty distribution at the intermediate milestones identified in Figure 10 for both the statistical dependence and the statistical independence cases.

5.1 Prior Analysis of Project Completion Time Uncertainty

At the start of the project, we conduct an analysis *a priori* using the example project network of Figure 4 and evaluate the project completion time distribution for the scenario where statistical dependence is applied (Figure 1) and the scenario where statistical independence is implied (see the introductory description of Section 5). We shall denote "DEP P" as the analysis scenario involving the BN model in Figure 1 (where statistical dependence is modeled), and also, denote "IND P" as the analysis scenario where statistical independence is implied among the marginal activity duration distributions. In Bayesian analysis terminology these distributions are referred to as prior distributions.

Figure 11 shows the project completion time probability density functions for both the "DEP P" and "IND P" scenarios. The step-like nature of each probability density function (pdf) displayed in Figure 11 is a result of the dynamic discretization methodology (Neil et al., (2007)) utilized in AgenaRisk®. The project completion time at 173.4 (from Figure 10b) is indicated in Figure 11 using a vertical dashed line. The "DEP P"pdf exhibits an expected value of 152.6 days compared to a value of 152.2 days for the "IND P" pdf. Thus the expected values for the project completion times evaluated under both scenarios exceed the CPM evaluated value of 144 days in Figure 10a. The "DEP P" pdf in Figure 11 exhibits a standard deviation of 10.8 days compared to 4.5 days for the "IND P" pdf, emphasizing the need for incorporating statistical dependence among activity durations in a prior uncertainty evaluation of project completion time.

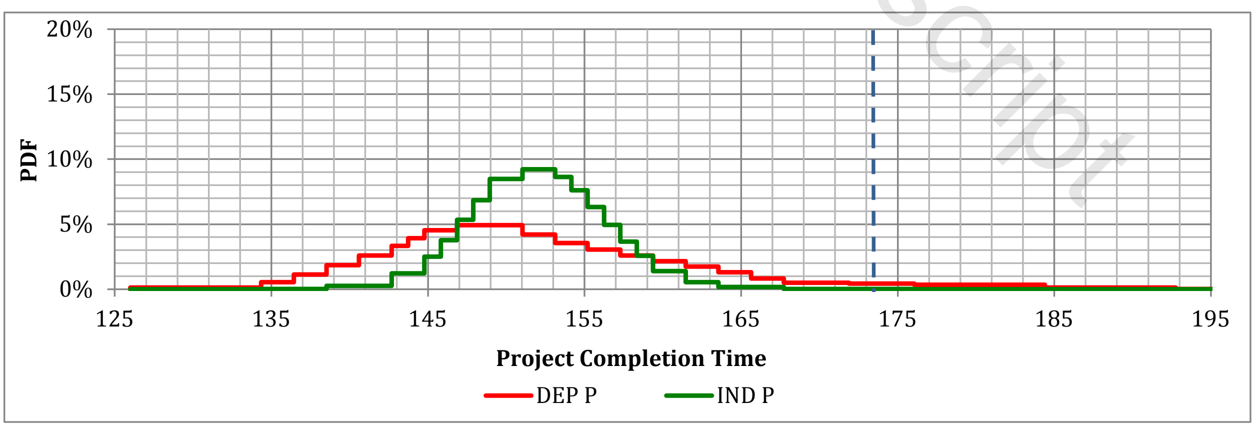


Figure 11: Prior project completion time distributions for the analysis scenarios "DEP P" and "IND P".

5.2 Posterior Analysis of Project Completion Time Uncertainty

To conduct analysis *a posteriori*, we shall use the traditional Bayesian notation "$\cdot|\cdot$". For example, the scenario description “DEP P|C” shall be used to designate the remaining project completion time uncertainty analysis involving the BN model in Figure 1 following the completion of Activity C. In Bayesian analysis terminology such a distribution is referred to as a posterior distribution.

Figure 12 displays the posterior distributions for the remaining completion time of the project at the four different milestones identified in Figure 10 (i.e. at completion of Activity C, Activity I, Activity F, and Activity K). The mean values and standard deviations for these project completion time distributions are listed in Table 3. Observe from Table 3 an increase of 10.7 days (=163.3 – 152.6) in expected value after only the completion of Activity C in the case of the dependence scenario. No change is observed in the case of the independence scenario.

Table 3: Mean values and standard deviations of remaining project completion time distributions evaluated at the project start and the four milestones identified in Figure 12 for the “DEP P” and “IND P” analysis scenarios

ANALYSIS SCENARIO	DEPENDENCE		INDEPENDENCE	
	Mean	St. Dev.	Mean	St. Dev.
P	152.6	10.8	152.2	4.5
P|C	163.3	8.4	152.2	4.5
P|CAEI	170.3	5.0	154.6	4.3
P|CAEIDF	171.2	4.0	157.0	3.9
P|CAEIDFGBJHK	172.6	2.1	164.8	2.6

Comparing Figure 11 with Figure 12a, a shift to the right is observed of the “DEP P” pdf and no shift of the “IND P” pdf. Following the completion of the first four activities C, A, E and I that have no predecessor, the difference in expected values for the dependence and independence scenario has increased from 0.4 day (=152.6 – 152.2) to 15.7 days (=170.3 – 154.6). One clearly observes from Figure 11, and Figure 12a to Figure 12b the reduction in uncertainty in the case of the dependence scenario. On the other hand, almost no reduction is observed for the independence scenario. This is confirmed by the first, second and third row in Table 3.

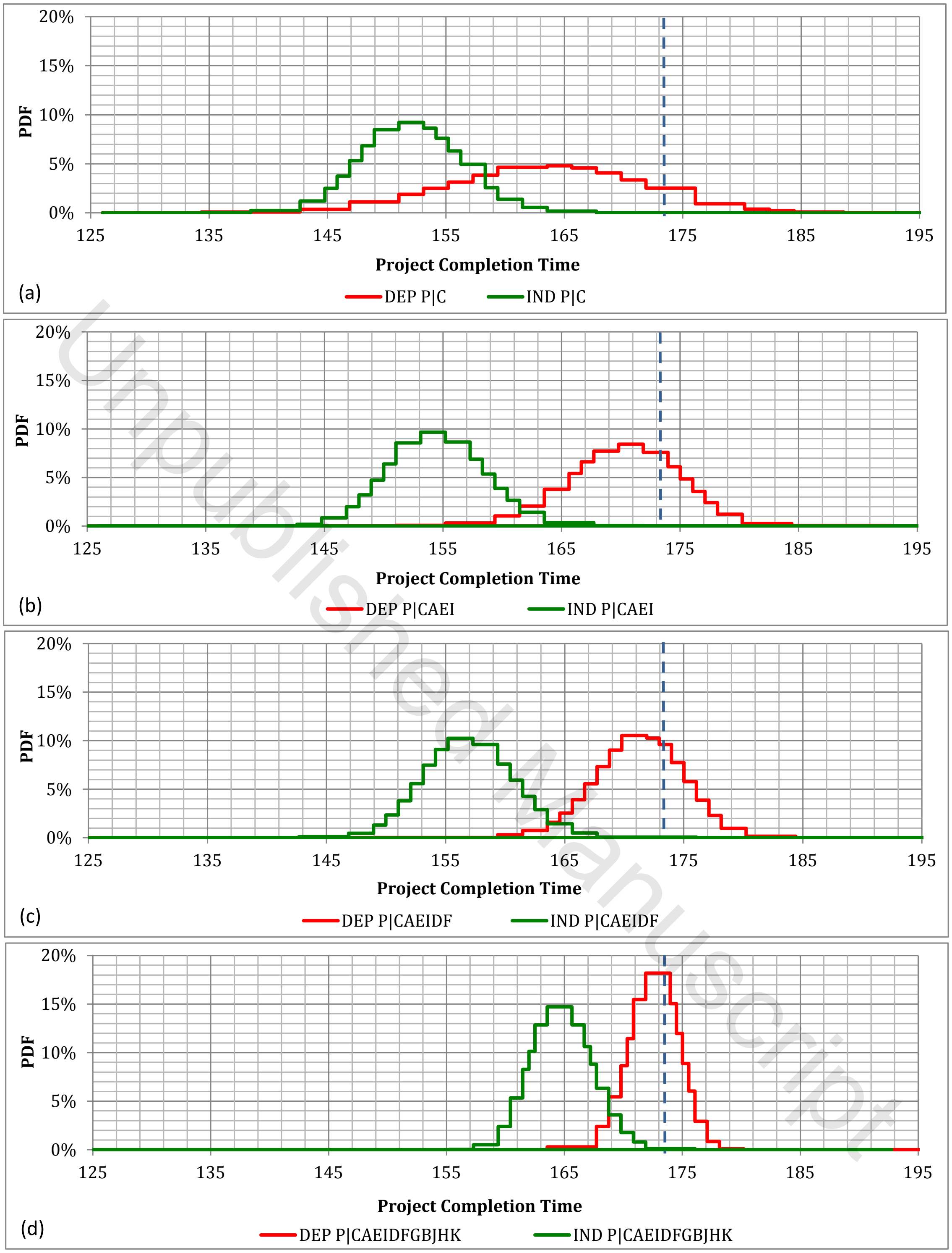


Figure 12: Posterior project completion time distributions for the analysis scenarios
(a) “DEP P|C” and “IND P|C” (b) “DEP P|CAEI” and “IND P|CAEI”
(c) “DEP P|CAEIDF” and “IND P|CAEIDF” (d) “DEP P|CAEIDFGBJHK” and “IND P|CAEIDFGBJHK”

At the third milestone (the completion of Activity F), about one third into the project (see Figure 10), the degree of uncertainty in both the dependence and independence scenarios (reflected by their respective posterior distributions) have about evened out (Figure 12c). From the fourth row in Table 3, a standard deviation of 4.0 days is evaluated for the dependence scenario (which means a reduction by 6.8 days compared to the prior standard deviation) and a standard deviation of 3.9 days for the independence scenario (which means only a reduction by 0.6 day). The former highlights the faster pace of learning about remaining completion time uncertainty, as the project progresses, when statistical dependence is incorporated amongst the activity durations in the project network using the BN Model in Figure 1.

Also, importantly at the third milestone, an expected completion time of 171.2 days is evaluated for the dependence scenario, whereas the expected value estimate in case of the independence scenario lags behind at 157.0 days. Recall from Figure 10b a completion time of 173.4 days was evaluated for the entire project using 95^{th} percentiles of activity durations for BN instantiation. Given that the third milestone happens about one third into the project this could leave sufficient time for project managers to take corrective action.

Finally, at the fourth milestone, about two thirds into the project, the independence scenario evaluation of the expected project completion time has caught up somewhat at 164.8 days compared to 172.6 days evaluated for the dependence scenario. It is important to note that at this point only one third of the project remains, perhaps too late for corrective action should one have adopted an analysis methodology that does not incorporate statistical dependence among the activity uncertainty distributions, and does not follow a BN approach towards monitoring remaining project completion time uncertainty. Observe from the last row in Table 3 that the standard deviation now is less for the dependence scenario (2.1 days) than the one evaluated for the independence scenario (2.6 days).

6 Summary and Conclusions

Using a Bayesian Network approach, we have demonstrated via an illustrative example the potential benefit of modeling statistical dependence among the activity durations in a project network on monitoring of remaining completion time uncertainty. Needless to say, one could argue that this potential benefit is overstated as all 18 activities in this illustrative example complete at the upper tail of their uncertainty ranges, which may be deemed an unlikely occurrence. On the other hand, the effect of positive statistical dependence among the activity durations does imply a higher likelihood of one activity being delayed given that others were delayed. We do believe this to be a more common occurrence in project networks than the case of negative dependence, where one activity being delayed would imply a lesser likelihood of another activity being delayed. As such, the BN model developed herein does not allow for incorporating negative dependence among activity durations.

Regardless, the example demonstrates the advantages of BN approach towards modeling statistical dependence among activity durations in terms of increased precision of the uncertainty analysis involved as well as in enhancements in the pace of learning about expected project completion time as the project progresses (as compared to an uncertainty analysis assuming statistical independence). Specifically, at the start of the project a larger variance is observed when dependence is incorporated; however, at about one-third into the project, in the example, the variance estimate is about the same in both the independence and dependence analysis scenarios. In the illustrative example used in the analysis, dependence levels among activity durations were set at an average correlation of about 0.5. In case of larger correlations, an even faster pace of learning may be observed. The implication of faster learning from a pragmatic point of view is that it may enable project managers to take corrective action sooner thereby potentially avoiding costly schedule slippage.

Acknowledgement

This work was supported, in part, by a fellowship from the Logistics Management Institute (LMI). We would also like to thank the editor and the referees whose comments improved earlier versions significantly.

7 References

Blomquist, N. (1950). On a measure of dependence between two random variables. *Annals of Mathematical Statistics, 21*, 593–600.

Cho, S. (2009). A linear Bayesian stochastic approximation to update project duration estimates. *European Journal of Operational Research, 196*, 585-593.

Cho, S., & Covaliu, Z. (2003). Sequential estimation and crashing PERT networks with statistical dependence. *International Journal of Industrial Engineering, 10*, 391-399.

Covaliu, Z., & Soyer, R. (1997). Bayesian Learning in project management networks. *American Statistical Association Proceedings, Section on Bayesian Statistical Science*, (pp. 257-260).

Fang, C., & Marle, F. (2012). A simulation-based risk network model for decision support in project risk management. *Decision Support Systems, 52*, 635-644.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association, 100*(470), 680-700.

Hahn, E. D. (2008). Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research, 188*(2), 450-459.

Herrerias-Velasco, J. M., Herrerias-Pleguezuelo, R., & van Dorp, J. R. (2011). Revisiting the PERT mean and variance. *European Journal of Operational Research, 210*, 448-451.

Jenzarli, A. (1994). *PERT belief networks, Report 535.* The University of Tampa, FL.

Khodakarami, V., Fenton, N., & Neil, M. (2007). Project Scheduling: Improved Approach to Incorporate Uncertainty Using Bayesian Networks. *Project Management Journal, 38*, 39-49.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53*(284), 814-861.

Malcolm, D., Roseboom, J., Clark, C., & Fazar, W. (1959). Application of a technique for research and development program application. *Operations Research, 7*(5), 646-669.

Neil, M., Tailor, M., & Marquez, D. (2007). Inference in Hybrid Bayesian Networks using Dynamic Discretization. *Statistics and Computing, 17*(3), 219-233.

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

The Standish Group. (1995). *The Chaos Report.* Boston: The Standish Group.

The Standish Group. (2009). *The Chaos Report.* Boston: The Standish Group.

van Dorp, J. R. (2005). Statistical Dependence through Common Risk Factors: With Applications in Uncertainty Analysis. *European Journal of Operational Research, 161*(1), 240-255.

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

van Dorp, J., & Duffey, M. (1999). Statistical dependence in risk analysis for project networks using Monte Carlo methods. *International Journal of Production Economics, 58*, 17-29.

Virto, M., Martin, J., & Insua, D. (2002). Approximate solutions of complex influence diagrams through MCMC methods. In S. Gamez (Ed.), *First European Workshop on Probabilistic Graphical Models*, (pp. 169-175).