Modeling Project Schedule Uncertainty due to a Programmatic Risk Factor using Bayesian Networks

by Ifechukwu C. Nduka

B.Eng. in Civil Engineering, September 2004, University of Nigeria, Nsukka
M.S. in Project Management, May 2008, The George Washington University

A Dissertation submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial satisfaction of the requirements
for the degree of Doctor of Philosophy

August 31, 2015

Dissertation directed by

Johan René van Dorp
Professor of Engineering Management and Systems Engineering & of Decision Sciences

The School of Engineering and Applied Science of The George Washington University certifies that Ifechukwu C. Nduka has passed the Final Examination for the degree of Doctor of Philosophy as of April 1, 2015. This is the final and approved form of the dissertation.

Modeling Project Schedule Uncertainty due to a Programmatic Risk Factor using Bayesian Networks

Ifechukwu C. Nduka

Dissertation Research Committee:

Johan René van Dorp, *Professor* of Engineering Management and Systems Engineering & of Decision Sciences, Dissertation Director

Thomas A. Mazzuchi, *Professor* of Engineering Management and Systems Engineering & of Decision Sciences, Committee Member

Michael Stankosky, *Research Professor* of Engineering Management and Systems Engineering, Committee Member

Royce Francis, *Assistant Professor* of Engineering Management and Systems Engineering, Committee Member

Ekundayo Shittu, *Assistant Professor* of Engineering Management and Systems Engineering, Committee Member

Refik Soyer, *Professor* of Decision Sciences and of Statistics, Committee Member

©Copyright 2015 by Ifechukwu C. Nduka
All rights reserved

Dedication

This dissertation is dedicated to the loving memory of

my dad and first hero,

Engr. Gabriel Ifechukwu Nduka.

Acknowledgement

The courage to logically challenge conventional wisdom; the joy of making a novel contribution to scientific knowledge; the opportunity to learn from and interact with some of the brightest minds; and very importantly, a phase in my life that introduced me to who I am at the core: an unwavering optimist. These were the hallmarks of my doctoral journey. I am thankful to God for the blessing of completing my doctorate, and also, thankful to everyone who encouraged me to make this a reality.

Specifically, I want to appreciate the intellectual guidance and support provided by all three of my doctoral mentors who individually served as my doctoral advisors (at different stages of my training as a doctoral student) – Dr J. René van Dorp, Dr Thomas Mazzuchi, and Dr Michael Stankosky:

- Dr J. René van Dorp: I want to thank you for all your support, intellectual guidance, and advice regarding my dissertation research. Very importantly, I appreciate the help and inspiration you provided which enabled me to effectively drive the course of my dissertation research. Thank you for bringing your expertise in applied probability and dependence modeling to bear in mentoring me on this research.
- Dr Thomas Mazzuchi: I want to thank you for all your support, intellectual guidance, and fatherly kindness. Very importantly, I am greatly obliged for the advice, encouragement, and freedom you gave me in choosing my dissertation research area: applications of Bayesian Networks to risk analysis. Thank you for your

mentorship and the knowledge I gained from you in probabilistic risk analysis. Your benevolence to me would remain evergreen in my heart.

- Dr Michael Stankosky: I am thankful for the unflinching encouragement, mentorship, and intellectual guidance I received from you. Very importantly, I am grateful for your mentorship in the field of knowledge management. Thank you for all your advice (both personal and academic). I appreciate all your kind support, and more so, I am glad to call you both a mentor and a friend.

I also want to thank other members of my dissertation committee: Dr Royce Francis – for the encouragement and advice he provided in exploring the Bayesian Network methodology; and Dr Ekundayo Shittu – for the valuable comments and feedback he provided which helped in writing this dissertation; and Dr Refik Soyer – for his expository questions as well as the insight learned from his published works in the application of the Bayesian paradigm to project management. Also, I appreciate the faculty of the GW Department of Decision Sciences – Dr Young Hoon Kwak, Dr Sanjay Jain, and Dr Denis Cioffi – who mentored me in project management.

The love and support of my family and friends were profound, and I would like to use this opportunity to express my heartfelt gratitude especially to: my mother, Ethel; my cousin, Dr Emeka Anyake; and my friend, Mr Sunny Ahuwanya. My gratitude also goes to the Logistics Management Institute (LMI) and the Department of Engineering Management and Systems Engineering for their financial support.

Above all, I am very thankful to God from whom all blessings flow.

A portion of this dissertation is part of the paper listed below; however, the content of this document is original for the purposes of this dissertation:

- Nduka IC, van Dorp JR. *Monitoring Uncertainty in Project Completion Times: A Bayesian Network Approach*. Submitted to the European Journal of Operational Research; Feb. 16 2015.

Abstract

Modeling Project Schedule Uncertainty due to a Programmatic Risk Factor using Bayesian Networks

The classical method of project risk analysis, for example, the Program Evaluation and Review Technique (PERT), ignores statistical dependence among project activities, which puts limits to its effectiveness as a robust method for probabilistic schedule analysis. Also, some programmatic risk factors have been identified as significant sources of uncertainty in project performance. However, given a project that is progressing within the project life cycle, monitoring of uncertainty in the project completion times resulting from a programmatic risk factor while taking advantage of activity duration statistical dependence, has not been addressed, to the best of our knowledge. In this study, we develop two methods that are based on Bayesian Networks (BN) for evaluating and monitoring uncertainty in project completion times, when an ongoing or progressing project is impacted by a programmatic risk factor. The BN methods developed in this study model statistical dependence in project networks using parametric relationships between nodes which reduce the burden of dependence specification in the BN model. In modeling the relationships of the variables of the BN models developed in this study, concerns about computational complexities and efficient modeling of interactions of variables of a project are enabled by the capacity of the specialized Bayesian Networks software (*AgenaRisk*®) used in the analysis. Other concerns of classical PERT such as: (i)

assumption of statistical independence is addressed by using conditional median as a measure of statistical dependence, and (ii) the constant PERT variance assumption is addressed using the Modified PERT Variance. Using the BN models developed in this study, we demonstrate that failure to incorporate statistical dependence grossly underestimates the total uncertainty in project completion times. The graphical dimension of our model, which benefits from the capacities of Bayesian Networks, gives more visibility about the model development and uncertainty analysis process, and which could be helpful to project analysts and managers by providing greater insight and formal mechanisms for interpreting how uncertainties in project performance measures emerge. More so, 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.

Table of Contents

Dedication iv

Acknowledgement v

Abstract viii

Table of Contents x

List of Figures xiii

List of Tables xvi

1 Introduction 1

1.1 Research Objective and Problem Statement 3

1.1.1 Assumption of statistical independence in simulation of project networks 3

1.1.2 Expert representation of uncertainty in project parameters and paucity of relevant data for effective project risk assessment 4

1.1.3 Analysis of complex interrelationships between project risk and project parameters 5

1.2 Bayesian Networks 7

1.3 Programmatic and Common Risk Factor 8

1.4 Research Methodology 10

1.5 Contributions 11

1.6 Organization of the document 12

2 Literature Review 13

2.1 Uncertainty in Project 15

2.1.1 Uncertainty in Project Scheduling 16

2.1.2 Program Evaluation and Review Technique (PERT) 17

2.1.3 Simulation of Project Activity Network 24

2.1.4 Critical Chain Project Management (CCPM) 26

2.1.5 Event Chain Methodology 28

2.1.6 Bayesian paradigm applied to management of projects 29

2.2 Bayesian Network approach for modeling project risk and uncertainty 30

2.2.1 The ERIC-S model 31

2.2.2 Modeling uncertainty in project schedule by mapping a CPM project network to a Bayesian Network 33
2.2.3 Methodology for project risk analysis by integrating Bayesian Network in a Monte Carlo simulation environment 35
2.3 Modeling statistical dependence and programmatic/common risk factors 37
2.3.1 Uncertainty analysis of projects and modeling of statistical dependence caused by common risk factors 38
2.3.2 Quantification of uncertainty in projects due to common risk factor (i.e. Requirement Change) 39
3 Bayesian Network Activity Dependence Model for Project Risk Analysis 41
3.1 Model framework 41
3.2 BN Model for *Method 1* 46
3.2.1 Common Quantile Level 46
3.2.2 Mode Level 46
3.2.3 Duration Level 48
3.3 AgenaRisk® representation of the BN model for *Method 1* 50
3.4 Activity Statistical Dependence 53
3.5 Eliciting Degree of Dependence 54
3.5.1 Eliciting Conditional Medians 54
3.5.2 Conditional Median Matrix 56
3.5.3 Average Conditional Median 57
3.6 Case Study 1 59
3.7 Prior Analysis 61
3.8 Posterior Analysis 66
3.8.1 Posterior Analysis of the Monitoring Variable (Common Quantile Level) 67
3.8.2 Posterior Analysis of the Activity Duration Level 68
3.8.3 Posterior Analysis of Project Completion Times 70
3.8.4 Posterior Analysis of Project Completion Times (given evidence at activity duration nodes) 75
3.9 Summary of *Method 1* 78
4 Bayesian Network Activity Dependence Model for Project Risk Analysis using Modified PERT Variance 81
4.1 Case Study 2 81

4.2 BN Model for Method 2 84
4.2.1 Common Quantile Level 84
4.2.2 Mode Level 85
4.2.3 Duration Level 86
4.3 AgenaRisk® representation of the BN model for *Method 2* 87
4.4 Statistical Dependence Specification 90
4.5 Eliciting Activity-to-Activity Conditional Medians 91
4.6 Uncertainty (*n*) and Dependence (β) parameter specification 94
4.7 Analysis of Case Study 2 101
4.8 Prior Analysis of Project Completion Time Uncertainty 104
4.9 Posterior Analysis: Monitoring Uncertainty 105
4.10 Summary of Method 2 109
5 Conclusions and Recommendation 113
5.1 Conclusions 113
5.2 Recommendation for Future Research 114
REFERENCES 115
APPENDICES 123
Appendix A: Bayesian Network Formalism 124
A.1 Background of Bayesian Approach to Uncertainty Analysis 124
A.2 Bayesian Network 126
A.3 Node Independence Assumption and D – Separation 129
A.4 Construction of Bayesian Networks 132
A.5 Bayesian Network Modeling with *AgenaRisk®* 134

List of Figures

Figure 1-1: Relationship between programmatic risk and other risk factors (INCOSE, 2006) 9

Figure 2-1: Distribution of all possible times for a project activity (Meredith & Mantel, 2006) ... 18

Figure 2-2: PERTBN model (Yong & Zhigang, 2011) 23

Figure 2-3: CCPM project schedule in comparison with the classic project schedule (Raz, Barnes, & Dvir, 2003)..... 27

Figure 2-4: BN model for activity duration uncertainty - Khodakarami et al. (2007)..... 34

Figure 2-5: BN for an activity with CPM parameters – Khodakarami et al. (2007)..... 34

Figure 2-6: BBN-MCS Framework for modeling project risk by integrating Bayesian Network within a Monte Carlo Simulation environment (Ordóñez Arízaga, 2007) 36

Figure 3-1: Example project for Case Study 1..... 43

Figure 3-2: Bayesian Network Activity Dependence Model for Project Risk Analysis for Method 1 (illustrated for the example project of Figure 3-1)..... 43

Figure 3-3: BN model for Method 1 implemented in AgenaRisk for the example project for Case Study 1 51

Figure 3-4: AoA representation of example project network 55

Figure 3-5: Project Completion Time CDF without Instantiation of Activity Duration Nodes 63

Figure 3-6: AgenaRisk® representation of the BN model for Method 1 (showing comparison with implementation for the Independence scenario) 65

Figure 3-7: Distribution of the Common Quantile Level (Monitoring Variable) after instantiation with: A) 5th, B) 50th, and C) 95th percentiles of activity duration values..... 67

Figure 3-8: Distributions of activity duration (given the instantiations of activities across the project) for: A) 5th, B) 50th, and C) 95th percentile values..... 69

Figure 3-9: Project completion time CDF (instantiated with 95th percentile values) for both the independent and dependent scenarios 72

Figure 3-10: Project completion time CDF (instantiated with 95th percentile values) for both the independent and dependent scenarios 73

Figure 3-11: Project completion time CDF (with instantiation values as 5th, 50th, and 95th percentiles) for both the independent and dependent scenarios..... 74

Figure 3-12: Distribution of Project Completion Times (given evidence of 5th percentile values of activity duration estimate across the project network) 75

Figure 3-13: Distribution of Project Completion Times (given evidence of 50th percentile values of activity duration estimate across the project network) 77

Figure 3-14: Distribution of Project Completion Times (given evidence of 50th percentile values of activity duration estimate across the project network) 78

Figure 3-15 Graphical summary of the procedure used in developing BN Model for Method 1 .. 80

Figure 4-1: Example project for Case Study 2 (Taggart, 1980) 82

Figure *4-2*: *Example project for Case Study 2 (represented as Activity-on-Node project network)* 83

Figure 4-3: Bayesian Network Activity Dependence Model for Project Risk Analysis for Method 2 (illustrated for the example project of Case Study 2) 84

Figure 4-4: BN model for *Method 2* implemented in AgenaRisk® for the *example project of Case Study 2 (representing the Dependence scenario)* 88

Figure 4-5: Example project network for Case Study 2 implemented in AgenaRisk (representing the Independence scenario) 89

Figure 4-6: Conditional median matrix showing path-dependent activity-to-activity conditional medians for the example project in Figure 4-2 93

Figure 4-7: (a) Average Prior Marginal Activity Variance as a function $\boldsymbol{\beta}$ and $\boldsymbol{n}$ and (b) Average Prior Activity-to-Activity Conditional Median as a function $\boldsymbol{\beta}$ and $\boldsymbol{n}$ 98

Figure 4-8: Data for the 3D graphs in Figure 4-7 99

Figure 4-9: Visual comparison of BN activity variance and PERT modified activity variance for the eighteen activities listed in Table 1 (for values n=3.5 and β=13) 100

Figure 4-10: P-P plots comparing BN marginal distributions with Least Squares Beta fitted distributions for (a) Duration Activity A (b) Duration Activity B 102

Figure 4-11: CPM evaluation of the example project in Figure 4-1 using - (a) activity duration most likely values shown in Table 4-1 (b) using 95th percentiles of activity durations shown in Table 4-2. 103

Figure 4-12: Prior project completion time distribution for the analysis of scenarios "DEP P" and "IND P" 105

Figure 4-13: 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" 107

Figure 4-14: Graphical summary of the procedure used in developing BN Model for Method 2 112

Figure A-1: Example BN model with three nodes 129

Figure A-2: Types of connection for defining conditional independence in a BN 131

List of Tables

Table 1-1: Sources of programmatic risk (Schutt, 1997) 8

Table 3-1: Conditional Median Matrix..... 57

Table 3-2: Case Study 1 input data and model parameters for β *= 40. (For prior sampling repeat Step 2 and sample from Beta's)* 59

Table 3-3: Case Study 1 input data and model parameters for β *= 40 showing high degree of dependence and Average Conditional Median,* $\overline{Med}(\cdot|\cdot) \approx 0.746$ 60

Table 3-4: Case Study 1 input data and model parameters for β *= 20. (For prior sampling repeat Step 2 and sample from Beta's)* 60

Table 3-5: Case Study 1 input data and model parameters for β *= 20 showing mild degree of dependence and Average Conditional Median,* $\overline{Med}(\cdot|\cdot) \approx 0.673$ 60

Table 3-6: Case Study 1 input data and model parameters for β *= 10. (For prior sampling repeat Step 2 and sample from Beta's).* 61

Table 3-7: Case Study 1 input data and model parameters for β *= 10 showing low degree of dependence and Average Conditional Median,* $\overline{Med}(\cdot|\cdot) \approx 0.604$ 61

Table 3-8: Percentile values for project activity durations 62

Table 4-1: Case Study 2 Input Data 83

Table 4-2: Parameters of Beta fitted marginal distributions of activity durations and their 95th percentiles 103

Table 4-3: Mean values and standard deviations of remaining project completion time distributions evaluated at the project start and the four milestones identified in Figure 4-11 for the "DEP P" and "IND P" analysis scenarios 108

Table 4-5: Comparison of PERT means with BN means of project activities for Case Study 2 111

1 Introduction

Projects and programs experience growing complexity partly as a result of many interacting variables that impact project performance measures of schedule, cost, and quality. Uncertainties inherent in the variables or parameters of a project have the capacity to shroud the level of impact these variables might have on a project, thereby causing ambiguity in predicted project performance. More so, given a project that is progressing within the project life cycle, monitoring of uncertainty in the project completion times resulting from impact of a programmatic risk factor while taking advantage of activity duration statistical dependence, has not been addressed, to the best of our knowledge. Dependence and interaction of variables in a problem domain play a significant role in risk analysis; but methods that integrate the quantification of uncertainty and analysis of interactions/dependencies of project variables and their corresponding effects on project performance are very limited in literature and practice (Khodakarami, Fenton, & Neil, 2007; Fang & Marle, 2011).

Programmatic risks factors may extend beyond operational sources of risk, and which can cause significant uncertainty on the predicted performance of projects and programs (Kindinger, 2004). These sources of programmatic risk comprise factors which have been identified as very critical to project performance (The Standish Group, 1995), and may include: material availability, requirement changes, communication problems, etc (Na, Li, Simpson, & Kim, 2004; Schutt, 1997; INCOSE, 2006).

This dissertation introduces two methods based on Bayesian Network (BN) – herein referred to as *Method 1* and *Method 2* – for modeling and evaluating dynamic uncertainties in project duration resulting from partial activity completions and statistical dependence among project activities influenced by a common risk factor. Statistical dependence among the marginal distributions of activity durations is modeled indirectly through the conditional relationships that exist between uncertainty nodes in the Bayesian Network – for example, in *Method 1*, activities' *mode relative distances* are modeled as Two-Sided Power (TSP) random variables with a common variance (and also with the shape parameter n unique for each project activity); and in *Method 2*, activities' *mode relative distances* are also modeled as Two-Sided Power random variables (but with the shape parameter n shared among all project activities). In addition, the modified PERT variance (Herrerias-Velasco, Herrerias-Pleguezuelo, & van Dorp, 2011) is used in the analysis involved in *Method 2*. For both methods discussed herein, activity uncertainty given its *most likely value* follow Beta distributions with a common parameter, reminiscent of classical PERT (Program Evaluation and Review Technique). The uncertainty model parameters are completely specified by directly eliciting lower and upper bound estimates for the activities next to the *most likely value* estimates. Dependence parameters are elicited indirectly through elicitation of probabilities that an activity completes above its median value given that another activity is already completed beyond its median (along the paths in the project network).

From a project monitoring standpoint, the model proposed herein provides a framework that illuminates the uncertainties in project schedule performance. More so, the overall scope and graphical dimension of the model provides greater insight in and formal mechanisms for updating evolving uncertainty in project durations as a project progresses. The proposed method provides a robust framework which incorporates complex interrelationships and dependencies among variables and parameters of a project network.

1.1 Research Objective and Problem Statement

The objective of this study is to develop a method based on Bayesian Networks for modeling duration uncertainty (both at the activity and project levels) due to a common (programmatic) risk factor, which addresses some inadequacies identified in techniques in project schedule analysis and the problem statements mentioned below. The problem statements and some of the inadequacies include:

1.1.1 Assumption of statistical independence in simulation of project networks

The assumption of statistical independence in the quantification of uncertainty of activities in a project network influenced by a common risk factor is specious and can result in the underestimation of total uncertainty in project duration (van Dorp & Duffey, 1999). It has been demonstrated that given a project network, the quantification of uncertainty in project duration cannot be completely defined by assuming statistical independence of distributions of individual activities (i.e. modeling individual marginal uncertainty distributions of activities in the project network); rather effective

quantification of uncertainty in project duration can be achieved when a complete multivariate distribution between uncertainties in activity duration for all project activities are indicated (van Dorp & Duffey, 1999).

The Monte Carlo Simulation (MCS) is a widely used risk-modeling application in industry. However, relevant literature in uncertainty analysis of project networks using MCS is indicative that the assumption of statistical independence implied by MCS is questionable. Application of MCS in uncertainty analysis of project networks has been identified to have some methodological limitations, thus: (1) It incorrectly assumes statistical independence for individual activities in Monte Carlo simulation of project networks, and hence does not address the issues of stochastic dependencies of individual project activities with common risk factors (van Dorp & Duffey, 1999; Duffey & van Dorp, 1998), and (2) It shows deficiencies in capturing feedback loops from chains of causality from programmatic risk sources (Willams, 2004).

1.1.2 Expert representation of uncertainty in project parameters and paucity of relevant data for effective project risk assessment

Projects are unique, one-off endeavors, and as such the uncertainty associated with project variables can best be considered as epistemic (i.e. uncertainty due to limited data and knowledge). Epistemic uncertainty as described by Bedford & Cooke (2001) can be quantified by experts, but cannot necessarily be measured. It is not uncommon that organizations involved in the use of standard risk management methods in managing project risks often have the difficulty to assess and manage project risks as a

result of paucity of relevant historical data to draw upon. Lack of relevant data for project risk assessment is even more pronounced when sources of project risk are novel; and in this instance, it becomes difficult or almost impossible to make effective analysis about the uncertainty in project schedule (or duration) parameters. Also, in many situations, enterprise factors may inhibit efficiency in system development efforts; and also some fundamental processes required to successfully drive a project or system development effort may be vague to the project team, and as such, elicitation of expert judgment may be required. It is important that a risk model encodes expert knowledge (Szwed, van Dorp, Merrick, Mazzuchi, & Singh, 2006), and the Bayesian paradigm in the form of Bayesian Network (BN) has the capacities to adequately model uncertainties and incorporate expert judgment. More so, appropriate analysis of uncertainty in projects should transcend variability and availability of data, to integrate structure and knowledge while also addressing ambiguity (Chapman & Ward, 2000). Also, reasoning about uncertainty can be formalized from subjective probabilities from expert judgment (Cooke, 1991), and proper managing of uncertainty in project should recognize the various sources of uncertainty and managing these sources in order to deal with desirable and undesirable consequences (Chapman & Ward, 2000).

1.1.3 Analysis of complex interrelationships between project risk and project parameters

Contrary to the dictates of the classical approach to risk assessment, uncertainty associated with risk is not a separate concept (Fenton & Neil, 2012). Every risk event (and all the resultant variables associated with the risk event) has uncertainty that is

characterized by the risk event's probability distribution. A major difficulty results in the classical approach to risk assessment when trying to address and incorporate various sources of uncertainty in relation to increasing project complexity; and in which BNs offer a compact and computationally tractable solution. A holistic approach to risk that incorporates a view of interconnected variables provides a robust method to risk assessment. It has been identified in some relevant literature that a comprehensive approach and holistic view of risk assessment should incorporate causes (or probabilistic influences) as well as consequences of risk events (Fang & Marle, 2011). The probabilistic and causal semantics as depicted by a Bayesian Network provide a structure for identifying, understanding, and quantifying complex interrelationships underlying risk events thereby revealing how risks emerge, are connected; and enable risk analysts make predictive and diagnostic reasoning about risk, as well as represent control and mitigations of risks (Fenton & Neil, 2012).

Project is a complex, dynamic system – and as a result, the complexity in a project arises from the interconnected and dynamic nature of variables that constitute a project, which causes resultant uncertainty and nonlinearity in project outcomes (Watt & Willey, 2005; Deshpande, 2012). More so, various sources of programmatic risk (e.g. organizational, environmental, human etc.) have the capacity to impact on project performance at any stage of the project life cycle, and these programmatic risk sources add complexity (and thus uncertainty in project performance) for which the risk assessment challenge then is to formalize an approach for estimation of uncertainty in project performance that has the capacities to integrate various project

variables/parameters and explicitly quantify uncertainty from different sources of programmatic risk.

1.2 Bayesian Networks

Bayesian Network (BN) is a probabilistic graphical model that captures the probabilistic (or causal) dependencies of a given set of variables in a problem domain through a directed acyclic graph. BN is recognized as a formalism and robust method for modeling and reasoning about uncertainty (Neil, Tailor, & Marques, 2007; Druzdzel & vanderGaag, 1995; Weber, Medina-Oliva, Simon, & Iung, 2012; Darwiche, 2009; Fenton & Neil, 2012; Pearl, 1988). It has the capacity to address the core limitations mentioned earlier by introducing into risk assessment process: a full representation of joint probability distribution over all variables; a dependency structure between variables that is motivated by probabilistic influence or causality; incorporating data and/or expert judgment. The combined capacities of BN imply that it can be conditioned on any state or subset of its variables, thereby supporting bidirectional forms of reasoning from *cause* to *effect* (and *effect* to *cause*). Modeling with Bayesian Network has received widespread appreciation in the academic and professional quarters, due mostly to BN's modularity, and mathematical and probabilistic elegance (Darwiche, 2009). Efficient and state-of-the-art algorithms implemented in *AgenaRisk*® (a specialized BN and simulation software for risk and decision analysis) enable risk-modeling that addresses the limitations mentioned above (Agena, 2012). Description of the Bayesian Network formalism and the utilization of *AgenaRisk*® to model uncertainty from a programmatic

risk factor are contained in the Appendix and subsequent chapters of this dissertation (i.e. Chapters 3 and 4).

1.3 Programmatic and Common Risk Factor

In this study, programmatic risk is defined as the type of risk produced by events that are beyond the control of the project or program manager (INCOSE, 2006; Schutt, 1997), which can impact on the outcome of a project or program. A programmatic risk factor is herein assumed as a source of dependence in projects, and thus is modeled as a common risk factor. The determination of the impact of *common risk factors* or *common causes* has found application in such areas as finance (Lustig, Roussanov, & Verdelhan, 2011; Viale, Kolari, & Fraser, 2009), uncertainty analysis (van Dorp J. R., 2005) organizational analysis (Trucco, Cagnoa, Ruggerib, & Grande, 2008), and fault tree analysis (Zhang, 1989).

Programmatic risks are generally classified by the source of factors that have the capacity to impact on the performance of a system or project. Some typical sources of programmatic risk are listed below in Table 1-1.

Table 1-1: Sources of programmatic risk **(Schutt, 1997)**

Source of Programmatic Risk	
Material Availability	Communication Problems
Personnel Availability	Labor Strikes
Personnel Skills	Requirement Changes
Safety	Political Advocacy
Regulatory Changes	Contract Stability
Environmental Impact	Funding Profile

A highly cited report on the statistics of failure of large-scale projects and complex system development projects is contained in *The Chaos Report* (The Standish Group, 1995, 2009). These reports on project failure statistics point to the fact that major causes of failure of large-scale projects are attributed to programmatic risk sources (The Standish Group, 1995, 2009; Al-Ahmad et al., 2009). Though some methods for modeling programmatic risk have been noted in literature, but much of these methods have methodological drawbacks given that they fail to model the dependencies and complex interactions between different programmatic risk sources (Fang & Marle, 2011).

The *Systems Engineering Handbook* (INCOSE, 2006) describes the context that exists between programmatic risk and other risk sources, here shown below as Figure 1-1 (as it relates to schedule and cost risks).

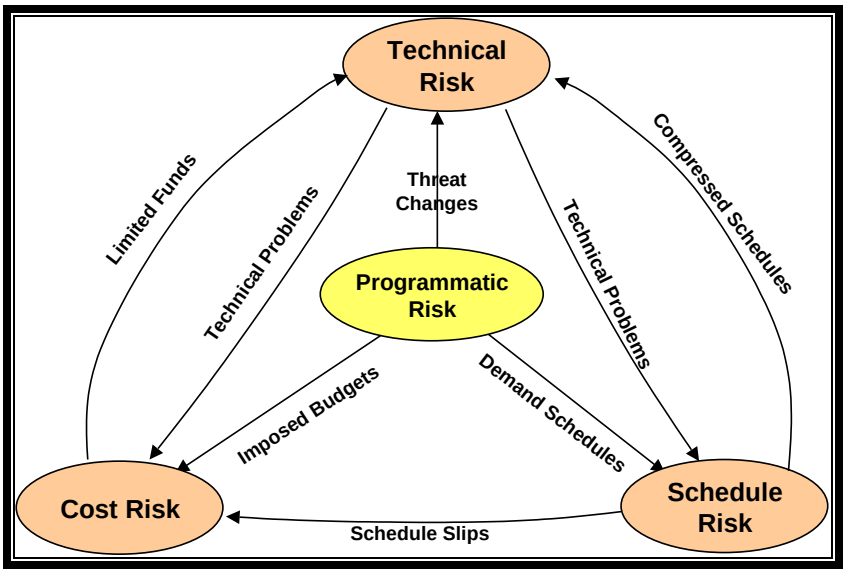


Figure 1-1: Relationship between programmatic risk and other risk factors **(INCOSE, 2006)**

1.4 Research Methodology

This study seeks to develop a method (given a project activity network) for determining uncertainty in project duration resulting from a programmatic risk factor using Bayesian Network. In general, the scope of this study is framed by the statement that: *Given a project activity network, the stochastic dependencies and uncertainty in project duration resulting from a programmatic risk factor can be adequately modeled using Bayesian Networks thereby providing an improved method in probabilistic schedule risk analysis.*

The methodology utilized in this study include: 1) Investigation of the current approaches in modeling and quantification of uncertainty in project activity networks, programmatic risk, and exposition of the seeming limitations therefrom; 2) Introduction of dependence modeling in the form of Bayesian Network, and description of Bayesian Network as a robust and improved method for modeling complexities and uncertainty in project parameters arising from a programmatic risk source; 3) Developing the topology of the Bayesian Network model which captures the dependencies and relationships of different variables of the problem domain being investigated in this study; 4) Determining dynamic uncertainties in project duration such that statistical dependence among project activities is realized by modeling their mode relative distances as Two-sided Power (TSP) random variable with a common variance, whereas activity uncertainty given its most likely distance is modeled as a Beta distribution with a common parameter; 5) Modeling and simulation of two example project networks using *AgenaRisk*® and application of each of the BN model developed herein to a case study of

project activity network to determine evolving uncertainty in project duration as a project progresses.

1.5 Contributions

In general, we have developed in this study a Bayesian Network model for modeling statistical dependence in project networks using parametric relationships between nodes which reduce the burden of dependence specification in the BN model. More so, the BN model developed herein has such capacities for evaluating and monitoring:

I. Uncertainty in project completion times when all the project activities are influenced by a common risk factor.
II. Evolving levels of uncertainty in project completion times as a project progresses.

Each of the two methods developed in this study (i.e. Methods 1 and 2) has both of the capacities mentioned above in (I) and (II). For Method 1, the BN model developed in this study is illustrated in Figure 3-2 and its implementation is shown in Figure 3-3. For Method 2, the BN model developed in this study is illustrated in Figure 4-3 and its implementation is shown in Figure 4-4.

Also, we have introduced in the BN model the use of the conditional median as a measure of expert representation of uncertainty in projects, which enable computation of statistical dependence in the project network as well as determining evolving uncertainties in project completion times.

The uncertainty in determining improved estimates of project completion times, which is often encountered in the execution phase of a project, can be illuminated by utilizing the model developed in this study thereby enabling project control managers to make more informed judgments about project performance. More so, the model's graphical dimension provides deeper insight and formal mechanisms for arriving at project decisions based on visible and auditable reasoning.

1.6 Organization of the document

This dissertation is structured into 5 chapters:

- Chapter 1 provides a general introduction of the research study, including the research objective and problem statements that drive the direction of the research. A brief discussion on Bayesian Networks is also presented. The concept and different sources of common risk factor and programmatic risk factor are discussed – with particular description given about requirement change as a programmatic (common) risk factor. An overview of the BN approach and contributions of this study are also described.
- Chapter 2 discusses insight from review of relevant literature.
- Chapters 3 and 4 provide detailed discussion of the BN models and methods developed in this study: *Method 1* and *Method 2* are discussed in details in Chapter 3 and Chapter 4 respectively.
- Chapter 5 presents the conclusion drawn from the research and discusses recommendations for future research.

2 Literature Review

The focus of this research study is concerned with uncertainty in project completion times particularly those caused by a programmatic factor (or common risk factor), for example, *Requirement Change*; and what can be done to effectively estimate these uncertainties in order to improve management decision-making. Bedford & Cooke (2001) asserted that any quantitative study of uncertainty must be backed by a full mathematical representation which comprises three things: *i*) axiomatic specification of the formal properties of uncertainty; *ii*) interpretations that connect the axiomatic specifications with observable phenomena; *iii*) measurement procedures that provide practical methods for interpretation of axiomatic specification. A consideration of the statements above is reflected in the approach utilized in this study as they provide a rigorous framework for describing uncertainty. First, our approach to quantification of uncertainty is expressed by the use of probability axioms; second, we utilize *subjectivist* interpretation of probability axioms which play a significant role in our approach to risk assessment as encapsulated in this study; and third, we use the notion of *degree of belief* as interpretations of probability axioms.

Within the context of this research study, our description of uncertainty concerns the outcomes of potential observations. Two general categorization of uncertainty appear in the literature – *aleatory* uncertainty and *epistemic* uncertainty (Bedford & Cooke, 2001; Hester, 2012). Aleatory uncertainty describes the inherent variability associated with a phenomenon in a system, while it is recognized that epistemic uncertainty is

characterized by lack of knowledge about specific characteristics of a system (Cogan, 2012).

Requirement change is a significant source of epistemic uncertainty in project performance. Studies in uncertainty modeling and risk analysis indicate the significance of making distinction between the two categories of uncertainty – aleatory and epistemic – as they provide basis for making clear modeling choices, as well as basis upon which uncertainty quantification can be done (Dubois & Guyonnet, 2011; Helton & Oberkampf, 2004; Bedford & Cooke, 2001). With respect to unavailability of relevant data for robust project risk assessment, the inherent uniqueness of projects presents some modeling challenges. Regardless of this implicit characteristic of a project with respect to data, nevertheless, decisions would be made and risk assessment would still have to be performed on the project parameter(s) of interest. Pate-Cornell (1996) explained that the *frequentist* view of probability is not applicable outside the domain of aleatory uncertainties, which, therefore, presents difficulties in the use of the *frequentist* approach to model uncertainties resulting from *requirement change* (a source of epistemic uncertainty). We argue the case about the use of the Bayesian Networks in propagating epistemic uncertainty resulting from a common risk factor (i.e. *requirement change*). We also describe that the uncertainty quantification method and the modeling approach explained herein should have the capacity to revise belief when new evidence is observed.

The comments above are indicative of the limited capacity of the *frequentist* approach to risk assessment, which offer minimal advantage to analyzing epistemic uncertainty involving unique events, or events involving incomplete information. This chapter describes a methodological review from existing literature regarding the aspects of project uncertainty, risk analysis of common risk factors, Bayesian Network modeling, and programmatic risk, which have relevance to the direction and focus of this research study.

2.1 Uncertainty in Project

In this study, project uncertainty is defined as a state of lack of knowledge about the true description of the existing state or future outcome (or a combination thereof) of a project variable or parameter. It is likened to an intangible measure of that which is unknown about any project variable or parameter.

Ward & Chapman (2003) and Ramgopal (2003) argue that current methods in project risk management place a limited focus on project uncertainty partly because the term "*risk*" has a more general connotation with threatening "*events*" than a description of sources of significant uncertainty. Managing uncertainty in projects should not be restricted to perceived events (i.e. threats and opportunities), but rather should also include a focus on identifying and managing sources of uncertainty which may affect the perception of potential threats and opportunities (Khodakarami, 2009).

A Guide to Project Management Body of Knowledge (PMBOK® Guide) – (PMI, 2008) holds the notion that uncertainty is the significant source of all project risk. However,

research trends in the academia and industry are indicative of the fact that quantification of uncertainty in projects is a big challenge as there are comparatively few tools or techniques which adequately deal with project uncertainty (Perminova, Gustafsson, & Wikstrom, 2008; Cleden, 2009; Asllani & Ettkin, 2007; Meyer, Loch, & Pich, 2002).

Our goal in this study is to model and quantify uncertainty in project duration due to a programmatic risk factor, but firstly, in the following sections, we provide further motivation for our proposed method (i.e. Bayesian Network) by describing various methods which appear in the literature for estimating project uncertainty.

2.1.1 Uncertainty in Project Scheduling

Project scheduling is a vital part of the overarching process of effective project management. Delivering projects on schedule can make a big difference to an organization in terms of profit, increased market share, enhanced competitive advantage, etc. Research suggests that in the development of project schedule during project planning, majority of practices are centered under the assumption that there is static and deterministic information about the scheduling problem and the project environment (Van de Vonder, Demeulemeester, & Herroelen, 2007). However, the reality is that projects are very dynamic with resulting uncertainties that may cause schedule disruptions.

A classical problem in project management lies in developing robust methods that effectively integrate uncertainty in project scheduling (Khodakarami et al., 2007),

though a number of project scheduling tools and methods have been developed over the past decades including the Critical Path Method (CPM), Program Evaluation and Review Technique (PERT), Critical Chain Project Management (CCPM), Event Chain Methodology, and Monte Carlo Simulation (MCS) of projects. Although CPM is restricted to deterministic project scheduling, it is regarded as one of the major project scheduling methods. Essentially the CPM assumes no resource constraints and concentrates in the generation of precedence relationships among project activities which is captured in a graphical structure. The focus of the CPM method is the identification of critical activities that are on the critical path so as to enable the concentration and application of resources in order to reduce the duration of a project. The graphical structure or project network generated using the CPM most often serve as the project schedule baseline. In Chapters 3 and 4 of this dissertation, project networks are used, in part, to demonstrate the modeling capacities of Bayesian Network in determining the uncertainty in project duration due to a programmatic risk factor.

2.1.2 Program Evaluation and Review Technique (PERT)

PERT was developed in the 1950s by the U.S. Navy as a tool for planning and scheduling of complex projects (Malcolm, Roseboom, Clark, & Fazar, 1959). Using probability theory PERT calculates the uncertainty in meeting schedule estimates of a project or project activity. It incorporates uncertainty in project scheduling even when there is incomplete information or sparse knowledge about the durations of the activities that make up a project. Using PERT it is also possible to estimate the duration that

corresponds to a project or project activity given a particular probability estimate (Callahan, Quackenbush, & Rowings, 1992). The major difference between PERT and CPM is that the CPM uses a deterministic approach in calculating project duration while PERT uses a probabilistic approach by incorporating uncertainty in activity duration.

The application of the classical PERT method follows some methodological steps that include: 1) building the project network where arcs represent activities and nodes represent events; 2) assigning duration estimates to each project activity: *optimistic* (*a*), *most likely* (*m*), and *pessimistic* (*b*); 3) calculating the expected value and standard deviation of duration times of each project activity. From a statistical viewpoint, *m* represents the mode while *a* and *b* represent the extremes of a hypothetical distribution (e.g. Figure 2-1) of duration times (Cottrell, 1999).

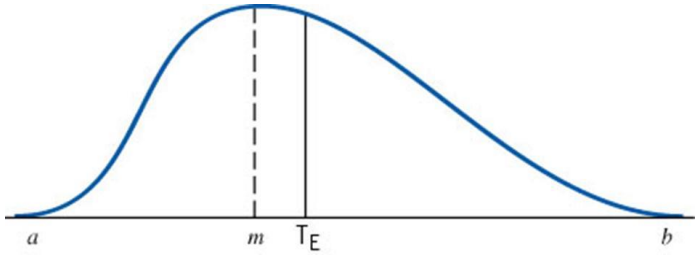


Figure 2-1: Distribution of all possible times for a project activity **(Meredith & Mantel, 2006)**

The expected time, T_E, is calculated by:

$$T_E = (a + 4m + b)/6 \qquad (2\text{-}1)$$

The measure of uncertainty for each project activity is calculated by the *variance*, σ^2, that is:

$$\sigma^2 = [(b-a)/6]^2 = [(b-a)^2/36] \quad (2\text{-}2)$$

Concerns about PERT has been largely discussed in the literature. Cottrell (1999) mentioned some major problems with PERT, which include: 1) Difficulty encountered by risk analysts or project schedulers in estimating accurately the parameters *a*, *m*, and *b* used in PERT. It was also pointed out by (Moder, Phillips, & Davis, 1983) and (Grubbs, 1962) that the PERT parameters of *a*, *m*, and *b* are based on subjective opinions and may not adequately reflect the statistical sampling of actual times; 2) Errors are prone to be introduced in the calculation of the mean and standard deviation of project duration as the mean and variance of activity duration which are calculated using equations (2-1) and (2-2) are actually estimates of the mean and standard deviation of a *Beta* distribution; 3) The classical PERT method generalizes all project activities to be applicable to the *Beta* distribution. It is noted in the literature that this generalization is fundamentally flawed as the *Beta* distribution is not representative of all project activities (Grubbs, 1962; MacCrimmon & Ryavec, 1964); 4) PERT considers only the critical path activities in its calculation of project duration probabilities; ignoring sub-critical activities and thereby causing a condition known as *merge event bias* (Callahan et al., 1992). This condition causes underestimation of duration and overestimation of variance (Halpin & Senior, 2010). More so, classical PERT assumes statistical independence of activities which makes it possible to calculate the project variance as the sum of variances of activities along a path. This assumption of statistical independence in PERT is inappropriate as studies have described the methodological

flaw in project risk analysis resulting from assumption of statistical independence (van Dorp & Duffey, 1999).

Numerous studies have been undertaken to improve the assumptions and expressions made in the classical PERT method, for example, the modified PERT variance (Herrerias-Velasco, Herrerias-Pleguezuelo, & van Dorp, 2011) provides an alternative to the limitations of classical PERT discussed by (Hahn, 2008) regarding the constant PERT variance assumption conditional on the range of support of an activity's duration. In the subsections below we review a few other studies that address some of the limitations of PERT.

2.1.2.1 Fuzzy PERT

Objections have appeared in the literature about some classical PERT assumptions like: the PERT methods for derivation of activity parameters of variance and mean; the use of *Beta* distribution (MacCrimmon & Ryavec, 1964; Cottrell, 1999). These criticisms about the assumptions used in classical PERT have motivated the emergence of the fuzzy PERT method as an alternative for the quantification of uncertainty related to activity duration. Bonnal, Gourc, & Lacoste (2004) posited that it is more appropriate to use the fuzzy set theory rather than the probability theory as axiomatic expressions for solving problems in project scheduling.

Fuzzy set theory has been in use in the past few decades, and it is gradually gaining acclaim as a method for modeling systems that are difficult to define precisely. The

fuzzy set is a set with elements that have a continuum of grades of membership (Zadeh, 1965). These elements constitute the fuzzy numbers and fuzzy intervals over the real line *R*. An attribute of the fuzzy set is a membership function which designates a membership grade between 0 and 1 to the elements of the set. The fuzzy PERT methodology incorporates imprecision and subjectivity into the entire modeling process; and instead of the use of probability distribution as in the classical PERT method, the fuzzy PERT method derives the project completion time by modeling activity duration times by membership functions based on possibility theory[1] and fuzzy set.

The methodological thrust of the fuzzy project scheduling technique is predicated on: the subjective nature of experts' estimates of activity duration times; lack of repeatability in activity and project realizations which, in some sense, conflicts with the notion of classical probability theory upon which classical PERT calculations are based (Chanas & Kamburowski, 1981). In essence, the proponents of this concept argue that the use of fuzzy variables to represent experts' opinion in a project network model is more robust in characterizing the nature of expert judgment and project performance measures than the probabilistic approach.

Since the first paper on the application of fuzzy set theory to project scheduling appeared in 1981 (Chanas & Kamburowski, 1981), many other publications about the applications of this concept to project scheduling have appeared in the literature.

[1] Possibility theory is a type of uncertainty theory and an extension of the fuzzy set concept (considered complementary to the probability theory), and which is used in analyzing and handling incomplete information.

Guiffrida & Nagi (1998) produced a chronological review of technical and research papers in the application of fuzzy set theory to various facets of project scheduling. Despite this growing trend in research publications; however, research indicates that given the problems involved in the derivation of membership functions for activity durations and the interpretation of results therefrom (Chanas & Kamburowski, 1981), the application of fuzzy set theory to project scheduling is regarded as more theoretical than practical. Further details of the technicalities involved in the application of fuzzy theory to PERT can be sourced from (Chanas & Kamburowski, 1981; Petrovan & Costea, 2010; Bonnal, Gourc, & Lacoste, 2004).

2.1.2.2 PERT and Bayesian Network

Mo & Zhao (2008) and Yong & Zhigang (2011) described methods that enable the formulation of PERT network as a Bayesian Network. In the frameworks described in their respective papers, BN was used to model dependencies between activity duration and project completion times.

The method used by Mo & Zhao (2008) involves approximating a stochastic PERT network by mixtures of Gaussian hybrid Bayesian Networks, and computing exact marginal distribution of project completion time using inference algorithms implemented in *Hugin*®, a commercially available Bayesian Network software package capable of handling Gaussian nodes. A major challenge identified with this approach is that complexity of inference is increased when there is arc reversal. Using this approach,

arc reversal, in essence, increases the domains of potentials making bidirectional inference very difficult.

The procedure introduced by Yong & Zhigang (2011) is represented by the PERTBN model shown below in Figure 2-2. In this framework, the duration *t(i), and ES*, *EF*, *LS*, *LF* of activity times are represented as nodes in the Bayesian Network. Conditional probability distribution of each node is described to represent the effect of immediate predecessor node(s) on the immediate successor node(s). Prior probability distribution is used to characterize the distribution of node(s) without immediate predecessor(s). They maintain that the use of this model (Fig. 2.2a) may create some complexity especially if a large project with many activities (nodes) is involved. As a result, the model shown in Fig. 2-2a may be converted to Fig. 2-2b to improve modeling efficiency and inference.

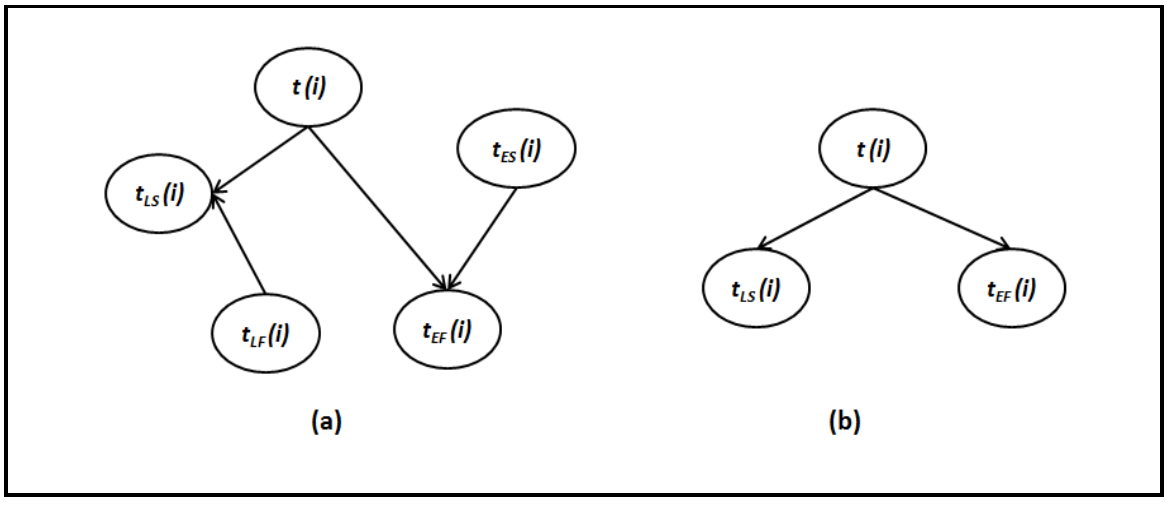


Figure 2-2: PERTBN model **(Yong & Zhigang, 2011)**

2.1.3 Simulation of Project Activity Network

Simulation of project network involves repeated sampling of random input values of project activity data and resultant calculation of associated project duration. The stochastic behavior of activities in a project network resulting from effect of a variable factor can be examined using simulation. Van Slyke (1963) introduced simulation as a technique for project network analysis; and since its introduction, simulation of project activity network as a method for project risk analysis has received widespread acceptance both in the academic and professional communities. The growing complexity, both in context and structure, which is normally evident in a project or project network, is a typical problem in project management that lends itself to simulation applications. Badiru (1991) described that at any given time, there are parts or sections of a project network that may be current and therefore available for direct observation and analysis; however, there are also significant sections of a project network that remain either in the past or are expected in the future, and which are readily analyzed using simulation.

Since its introduction, the use simulation for project management purposes has undergone several evolutionary changes and refinement, and recently many simulation-based software applications have been developed to run simulation of project networks in an efficient manner. However, issues related to methodological validity – including stochastic risk dependencies between project activities, elicitation of expert judgment for uncertainties of activity duration – remain sources of concern (Duffey & van Dorp, 1998).

2.1.3.1 Monte-Carlo simulation and PERT

Extensive research has been conducted in past few decades regarding the use of Monte Carlo methods to model project uncertainty (Badiru, 1991; Van Slyke, 1963; Khedr, 2006). The use of Monte-Carlo simulation approach more accurately solves PERT networks, as well as produces estimates of quantities for project planning that are improvements of the assumptions used in the classical PERT method (Finley & Fisher, 1994). The case of *merge event bias* which characterizes the classical PERT method is adequately handled with the use of Monte-Carlo simulation thereby producing unbiased estimates of mean and variance of project duration as well as the distribution of project duration (Douglas, 1978). Also, Monte-Carlo simulation of project networks gives a measure of the criticality (a probabilistic estimate of an activity being on the critical path) of an activity. The criticality index of activities provides a standard to determine which activities to focus on thereby improving project related decision made by management.

Effective modeling of project uncertainty is a major topic in Operations Research, and though many applications of Monte Carlo method have been introduced to address project uncertainty, however, the shortcoming of this method in accurately simulating correlations between activities as identified by Finley & Fisher (1994) remains one of the concerns with the use of Monte Carlo methods for risk assessment of project schedule.

2.1.4 Critical Chain Project Management (CCPM)

The Critical Chain Project Management (CCPM) is an innovative method for managing projects which incorporates the concepts of systems thinking and Theory of Constraints (TOC) to address uncertainty and variation in the duration of project activities (Leach, 1999). Developed by Goldratt (1997), it has some significant differences from the classical methods of project planning and control such as the CPM approach.

CCPM states that the likelihood of completion for duration estimates is 95%, and therefore suggests a reduction to 50% likelihood of completion for duration estimates. The difference between the 50% likelihood of completion and 95% likelihood of completion, called a buffer, is aggregated at the end of activity chains. CCPM suggests that the aggregated project buffer is less than the sum of all the individual margins or slack built in to each project activity thereby making allowances for uncertainty of estimates and project performance. Figure 2.3 shows a schematic description of CCPM method in comparison with the classic method of project scheduling.

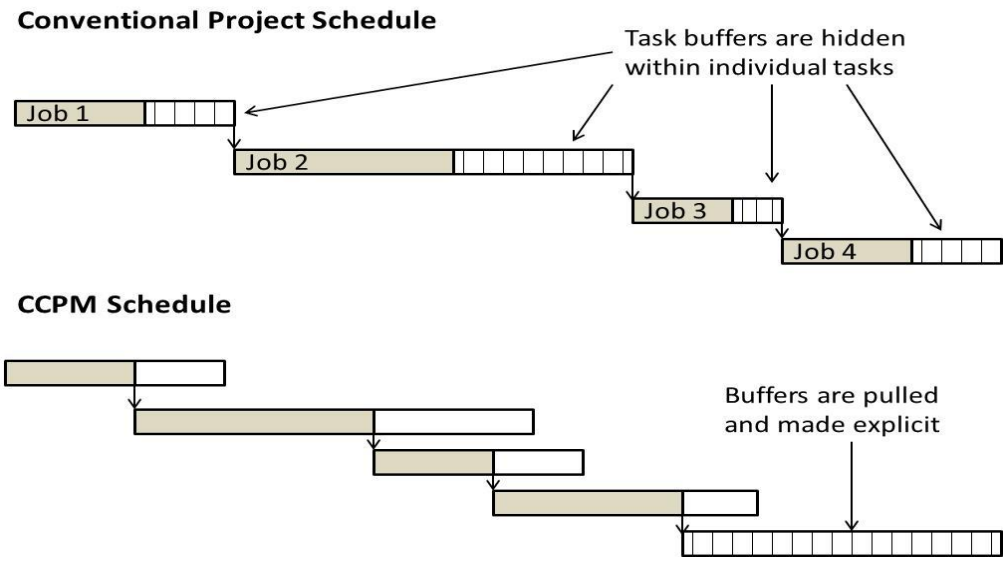


Figure 2-3: CCPM project schedule in comparison with the classic project schedule **(Raz, Barnes, & Dvir, 2003)**

CCPM follows a process that first identifies project activities and their respective duration estimates. This follows the development of a project network with due consideration given to availability of resources and activity dependencies. The critical chain is identified as the set of precedence and resource-dependent activities that results in the longest path to project completion, after resource leveling and assuming that resources are finite. If resources are assumed as infinite, then the critical chain of a project will be the same as its critical path. The next steps in the CCPM planning process include recalculation of the project schedule duration estimate by shortening the duration estimates of individual activities that make up a project. Raz, Barnes, & Dvir (2003) mentioned that the logical underpinning used in CCPM for shortening the activity duration estimates is based on the understanding that: activities in a project are prone to some levels of uncertainty; there is always over-estimation of project activity

durations; the slack built into activity duration are most often not utilized as tasks may be completed ahead of schedule.

Some criticisms about CCPM have been identified in the literature, which include: 1) CCPM is built under the assumption of statistical independence. This assumption is highly suspect as it has been demonstrated by van Dorp & Duffey (1999) that modeling marginal uncertainties of activities in a project network underestimates the total uncertainty of a project; 2) CCPM typically focuses on uncertainty inherent in a project schedule, but does not accommodate for analysis of probabilistic influence (or causal analysis) of the sources of uncertainty in a project schedule.

2.1.5 Event Chain Methodology

Event Chain Methodology (ECM) is an uncertainty modeling and schedule network analysis technique recently developed by researchers at the Intaver Institute. It leverages the strengths in existing methods and is considered as an advancement over the classic methods of Program Evaluation and Review Technique (PERT), Critical Path Method (CPM), Monte Carlo simulation method, and the Critical Chain Project Management method (Virine & Trumper, 2007).

The ECM is founded on the notion that in certain circumstances there are events that will alter the plan of a project. Resulting from the foregoing comment, ECM proposes a novel method in project scheduling with uncertainty by identifying and managing events or event chains that can potentially impact projects. The methodological underpinning

of the ECM consists of processes that model and visualize the impact of external event(s) on project activities (Intaver Institute Inc., 2005) by:

- Describing of the actual moment of an event and how the impact of an event can cause *excitation* – the process by which an activity changes its state as a result of impact from an event.
- Describing the ripple effects on project activities as a result of impacts of series events.
- Describing the relationships and correlation among events, as well as quantifying the cumulative effect of events and event chain on project schedule by means of simulation.

However, the uniqueness of projects makes historical data about project parameters rare, and therefore we argue that the effectiveness of ECM is limited in its use for modeling project uncertainty given that the methodology relies on historical data for calculating probability of events.

2.1.6 Bayesian paradigm applied to management of projects

Methodological improvements to some of the limitations of PERT using the Bayesian paradigm are well documented in literature. Noteworthy is the method proposed by Covaliu & Soyer (1996) in which the management of projects is designed as a sequential decision problem under uncertainty. Using the three models they proposed which represent three different contexts in project management – resource allocation model, common environment/resource model, and resource sharing model – they described

that at each stage of a project, decision regarding which project activity to start or crash is conditional on changes in resources and activity completion times. Using the Bayesian paradigm their proposed method updates the probability distributions of upcoming project activities given that a particular project activity has been completed. Dependence among project activities and resources are modeled as a multivariate distribution and represented in the form of an influence diagram. However, they mention that it becomes computationally intractable to apply their model to complex projects, or simply any project whose activities are neither in parallel nor consisting of more than 2 or 3 activities. Extension of the common environment/resource model is also described in Covaliu & Soyer (1997) in which the Gibbs sampler and rejection sampling are implemented in their proposed model to enable updating of parameter distributions and subsequent computation of predictive distributions of upcoming activity durations in a project network. This extension of the common environment/resource model enables Bayesian learning in project networks through the use of Markov Chain Monte Carlo (MCMC) methods.

2.2 Bayesian Network approach for modeling project risk and uncertainty

Bayesian Network as a tool for reasoning under uncertainty have been deployed to a wide range of fields and domains ranging from system reliability to medical diagnosis, but despite the growing trend in the application of BN across many fields, there remains very limited literature that describes the application of BN to project uncertainty and project risk management. This section reviews some of the most representative findings

about models and frameworks available in the literature that applies BN to project uncertainty and risk management.

2.2.1 The ERIC-S model

The ERIC-S model is regarded as the first building construction schedule risk model noted in literature that applies the Bayesian Network approach to estimate probabilistic values of activity duration based on some project characteristics (Nasir, McCabe, & Hartono, 2003). Essentially the model provides upper and lower activity duration limits based on the unique characteristics of a project, and which are used in stochastic schedule in Monte Carlo simulation or PERT. Nasir et al. (2003) described 4 major steps used in developing the ERIC-S model, and of which an overview is mentioned below:

1) Identification of variables – This involves a listing of factors or variables that are particular to the problem domain. The variables which are of two types – schedule risk variables and activity variables – indicate the scope of the work and the extent to which details of the problem domain are captured by the model. The risk variables capture various conditions of the project which are entered as evidence in the input node and activity variables serve as output nodes that describe activity duration estimates that are entered in the Monte Carlo simulation.

2) Identification and modeling of relationships – The conditional dependence relationship between the identified variables was performed by means of expert judgment elicitation and modeling of variables was performed using Bayesian Network approach. 9 experts were used in this stage of the elicitation process to determine the

conditional relationships between variables, with each expert using a prescribed ranking: *0 = no relationship; 1 = weak relationship; 2 = strong relationship; and 3 = very strong relationship.* Based on a combination of input of experts, experience of developers of the ERIC-S model, and literature review, relationships between variables that were identified as extremely weak were not considered as a basis for *cause-effect* relationship. In modeling the relationships between variables, a maximum limit of 4 parents were set for each variable in order to control the size of node probabilities.

3) Quantification of relationship between variables – This was also performed using expert judgment elicitation as there were insufficient data available. Given the large amount of quantification of relationships required for the model (i.e. 476 probabilities for the activity nodes and 366 probabilities for the risk nodes), the questionnaire used for the elicitation was designed in sets to give a logical distribution to questions that corresponds to each variable, thereby enabling the experts to give coherent responses without loss of rationality in the relationships of the variables. Responses from each of the 7 experts used in this stage of the elicitation process were both qualitative and quantitative, with each expert using up more than one session of 40 minutes per session. Three or more expert viewpoints were sought for each relationship between variables, and average values of these viewpoints were used as input to the model. In cases where the upper and lower limits of the expressed viewpoints of experts differ by 30%, a review of expert responses was required.

4) Model verification – In determining the efficiency and effectiveness of the ERIC-S model, the model was applied to some projects in which the following processes were utilized: a) Six risk nodes (variables) were randomly selected and sensitivity analysis was performed to determine the effect of changes in the values of these variables to a project. b) Also, the predictive capacity of the ERIC-S model was tested using several project cases drawn from 13 different projects.

The ERIC-S model was validated by the result of the tests conducted using the processes mentioned earlier. However, a major limitation of the model is that it is considered very restrictive given that it is applicable only to building construction projects in Canada, as the parameters of the model reflect only condition obtainable in Southern Ontario, Canada.

2.2.2 Modeling uncertainty in project schedule by mapping a CPM project network to a Bayesian Network

Khodakarami et al. (2007) developed a BN framework that models sources of uncertainty and their effects on duration of project activity. The framework called the Bayesian Critical Path Method (BCPM) describes an approach for determining schedule uncertainty in which CPM parameters of *Earliest Start (ES)*, *Earliest Finish (EF)*, *Latest Start (LS)*, and *Latest Finish (LF)* are incorporated into a Bayesian Network. Figures 2-4 and 2-5 are the schematic description of the BNs developed by Khodakarami et al. (2007) to model the influence of different sources of uncertainty on the duration of a project activity.

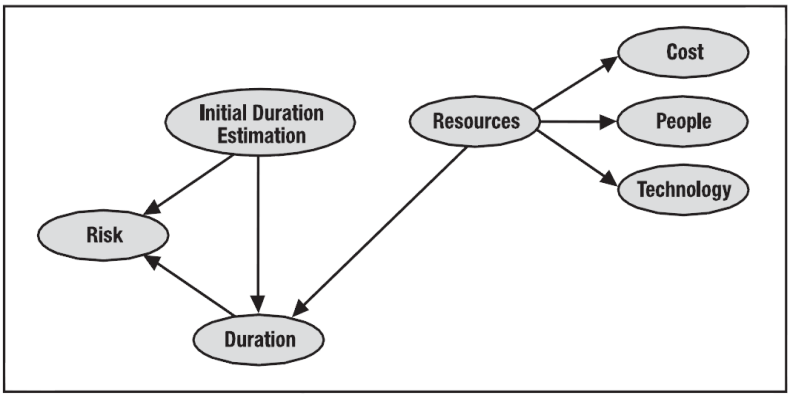


Figure 2-4: BN model for activity duration uncertainty - Khodakarami et al. **(2007)**

They reasoned that in order to map a CPM network to a Bayesian Network, it is important to first map the CPM parameters of a single activity (Fig. 2-4), where CPM parameters (*ES, EF, LS*, and *LF*) of a project activity are represented as variables of a Bayesian Network. Each activity in a CPM network is mapped using the BN model shown in Fig. 2-3, which has nodes that are representative of the parameters of an activity.

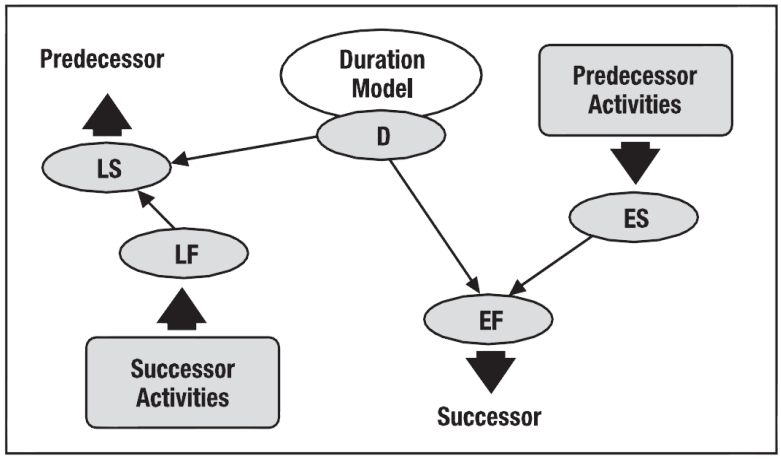


Figure 2-5: Bayesian Network for an activity with CPM parameters – Khodakarami et al. **(2007)**

Essentially, the BN models described above work in conjunction with each other to characterize uncertainty in project schedule using the following procedural steps: a)

mapping the duration uncertainty BN model (Fig. 2-4) to a project activity. In this way different sources of uncertainty that can impact on activity duration are captured and the uncertainty distribution estimated, and b) mapping Fig. 2-4 to the duration node, *D,* in Fig. 2-5 and computing the CPM estimates for the *forward pass* and *backward pass.*

The BCPM framework shows considerable promise in modeling uncertainty in project networks; however, when large project networks are involved the computational complexity of propagating inference across such large project networks becomes difficult or, in some cases, impracticable.

2.2.3 Methodology for project risk analysis by integrating Bayesian Network in a Monte Carlo simulation environment

A risk analysis methodology developed by Ordóñez Arízaga (2007) integrates schedule and cost uncertainties in projects by incorporating the effects of probabilistic correlation of work package cost and duration. He argues that the use of correlation matrix as input parameters in current project risk analysis approaches can be computationally cumbersome as the number of project variables increase, because of the corresponding combinatorial increase in the number of correlation coefficients. His research proposed a methodology – the *BBN-MCS* model – that integrates Bayesian Network within a cost-schedule Monte Carlo simulation model.

Ordóñez Arízaga (2007) reasoned that BNs are adequately suitable for project risk analysis and modeling of dependency and non-additive impacts due to simultaneous realization of different risk factors. In this methodology, the Monte Carlo simulation was

used in modeling the occurrence of independent risk, which is then propagated across the Bayesian Network to determine the posterior probabilities of dependencies among risks and the resultant impacts on cost and schedule. Figure 2-6 is a schematic description of the BBN-MCS model and the processes involved in project risk analysis by integrating BN in a Monte Carlo simulation environment.

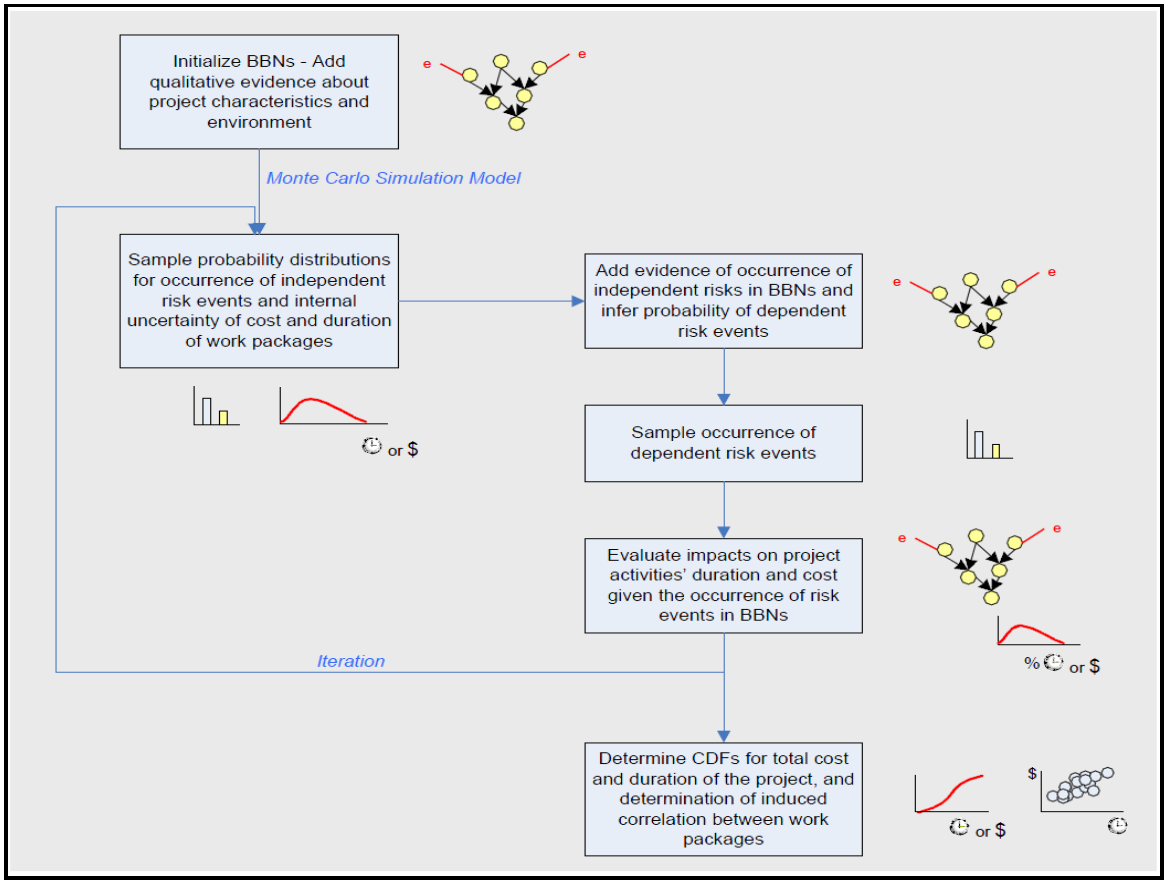


Figure 2-6: BBN-MCS Framework for modeling project risk by integrating Bayesian Network within a Monte Carlo Simulation environment **(Ordóñez Arízaga, 2007)**

The BBN-MCS model incorporates both the strengths inherent in the modeling capacities of Bayesian Network and Monte Carlo simulation methodologies. However, the author identified some methodological gaps in the model given that it does not: i)

explicitly identify any formal approach for eliciting node probabilities for the Bayesian Network used in the model; ii) the model is data-dependent, and given the novelty that is characteristic of projects, data may not be readily available, and hence the model may not be useful in this instance.

2.3 Modeling statistical dependence and programmatic/common risk factors

Uncertainty resulting from programmatic risk sources is a rich topic in existing academic research; however, some fundamental challenges have been identified as critical to the effective analysis of programmatic risk. Kindinger (2004) identified some of these major challenges in measuring programmatic risk to include: 1) *Completeness* – Sources of programmatic risk may extend across the entire gamut of a project or enterprise thereby posing challenges for analysis and modeling; 2) *Independence* – an assumption that is often applied in risk analysis is that risk factors may be normally distributed and independent thereby allowing for summation of their means and variances; 3) *Integration of varied sources* – Sources and impacts of programmatic risk are enmeshed in a web of relationship with other project parameters or enterprise factors, thereby posing a challenge for effective analysis and interpretation ; 4) *Data availability* - the uniqueness of projects often make relevant data for effective risk analysis very difficult to come by.

Given the challenges mentioned above as well as the varied nature of sources of programmatic risk, there is no general method for analysis and modeling of programmatic risk. Analysis of programmatic risk is performed based on the peculiar

nature of the programmatic risk factor and the environment upon which such programmatic risk factor may have an impact. Reminiscent of the factor model used in finance, *programmatic or common risk factor* in the context used in this study is reflective of the terms – *systematic* and *non-systematic* risks – used in financial modeling.

Other recurring methods that have been proposed in the literature include: Multi-Criteria Decision Analysis; Discrete Event Simulation; and Bayesian Network Modeling. Also, the Advanced Programmatic Risk Analysis and Management model (APRAM), which was developed to model risk of failure of dependent projects, has been applied to various scenarios (including aerospace and construction projects) to measure and analyze some programmatic risk factors (Dillon, Pate-Cornell, & Guikema, 2003).

2.3.1 Uncertainty analysis of projects and modeling of statistical dependence caused by common risk factors

A method was developed by van Dorp (2005) for building multivariate dependence between variables by identifying common risk factors as a source of dependence. The method leverages the diagonal band (DB) copula approach thereby allowing for efficient use of expert judgment to specify dependence parameters. The joint distribution generated by application of this method to PERT analysis has been demonstrated to be useful in assessment of uncertainty in project completion times.

Cho (2009) developed a hierarchical linear Bayesian estimation model in which statistical dependence among activities of a project is enabled between the activity

duration and amount of resource, and also between a risk factor and activity resource amount. It was described that the linear Bayesian approximation method provides the same computational benefit when compared to the full Bayesian approach. However, a major limitation of this method is in its assumptions and use of the joint normal distribution in modeling random variables – in the absence of the normality assumption, the usefulness of the method needs to be evaluated considerably.

2.3.2 Quantification of uncertainty in projects due to common risk factor (i.e. Requirement Change)

Requirements form the foundation of a project (INCOSE, 2006). It has been defined as that condition that must be met by a system or project in order to satisfy a contract, specification, or standard (PMI, 2008). However, due to complexity of a project environment and dynamics of systems engineering effort, requirements often change over the course of a system or project life cycle thereby causing resultant impacts on project parameters.

Methods for quantifying and modeling *Requirement Change* (also referred to as *Engineering Change Orders*) and its impacts on project performance have emerged over the few decades. Kulk & Verhoef (2008) developed a mathematical model that takes into account the size and duration of projects to determine the maximal requirement volatility tolerance for a project or portfolio of projects. A study by Wang, Li, Wang, Zhang, & Wang (2012) investigated the effect caused when the modification of a particular requirement causes corresponding changes in dependent requirements. They

argued that impact analysis from a scenario such as the aforementioned can be complex as a result of the ripple effect and dependency relationships in the requirements. A simulation approach – DepRVSim (Requirement Volatility Simulation considering Dependency relationship) – was proposed by Wang et al. (2012) on the consideration that dependency relationships among requirements must be considered whenever impact analysis of requirement change on project performance is evaluated.

3 Bayesian Network Activity Dependence Model for Project Risk Analysis

In *Appendix A* of this dissertation we describe an overview of the methodological underpinning of Bayesian Networks, as well as the unique features of *AgenaRisk*® which enable the BN modeling described in this Chapter. Also, in Section 1.3 we provided an overview of programmatic risk factors and explained that a programmatic risk factor is a source of dependence in a project activity network and thus can be modeled as a common risk factor. In this chapter we introduce the methodological framework used in developing the BN models described in this study.

3.1 Model framework

The *Bayesian Network Activity Dependence Model for Project Risk Analysis* which is introduced in this chapter describes the methodological framework of the two BN models developed in this study. The BN models developed in this study extend and build on two concepts: 1) The BCPM model introduced by (Khodakarami et al., 2007) in which uncertainty in project schedule is modeled by integrating BN and CPM methods; and 2) The Statistical Dependence modeling paradigm as it applies to project networks (van Dorp & Duffey, 1999; van Dorp J. R., 2005).

The dynamic nature of projects oftentimes result in the inevitability of changes during the project life cycle, and *common risk factors* have been identified as major causes of changes in a project and thus having high capacity to impact on project performance

(Zowghi & Nurmuliani, 2002). Incorporating uncertainty in project scheduling and specifying dependence resulting from common risk factors among project activities, remain a classical problem in project scheduling. Khodakarami et al. (2007) argue that most of the methods for handling risk and uncertainty in project scheduling are very limited given that they are event-oriented and tend to model just the impact of possible threats on project performance. Also, they maintain that the classical methods ignore the inherent sources of uncertainty and interaction of risk factors with project parameters.

The BN models developed in this study enables monitoring and determining of project completion time uncertainty, and also incorporates statistical dependence among the marginal distributions of activity durations in a project network. In Figure 3-1 we show an example project network, which hereafter would be used for analysis in Case Study 1. In Figure 3-2 the BN model is illustrated with the example project network representing a 5-activity project network (hereafter referred to as **Case Study 1**). As illustrated in Figure 3-2, the graphical layout of the BN model consists of three levels: the Common Quantile Level; the Mode Level; and the Duration Level.

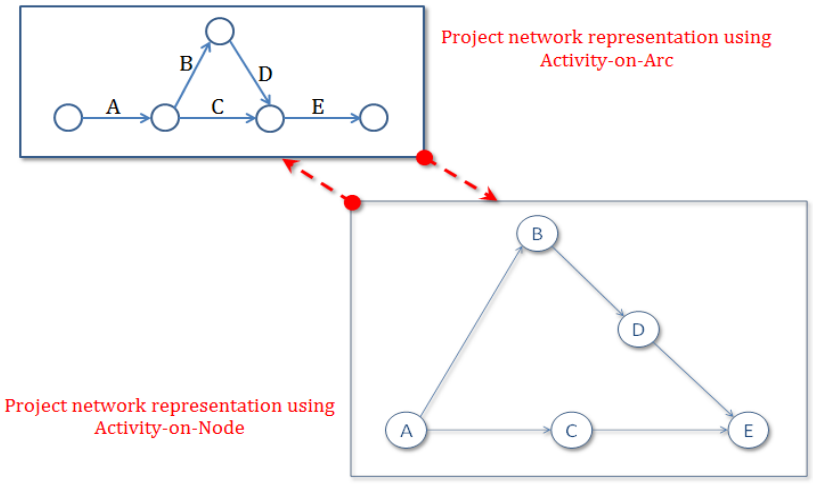


Figure 3-1: Example project for Case Study 1

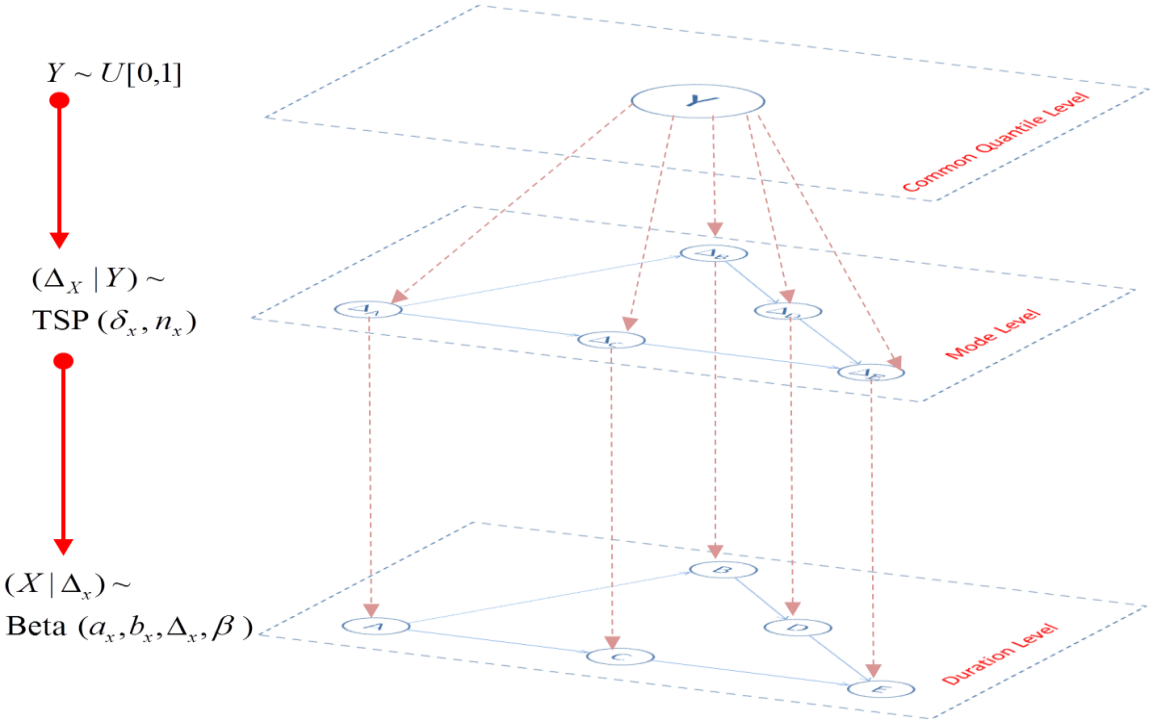


Figure 3-2: Bayesian Network Activity Dependence Model for Project Risk Analysis for Method 1 (illustrated for the example project of Figure 3-1)

We give a general classification to the methods used in developing the BN models in this study:

- ***Method 1***: This describes dependence specification using the mode relative distance uncertainty, which is discussed as *Case Study 1* in Chapter 3 using an example 5-activity project network shown in Figure 3-1. The unique characteristics of this method require elicitation procedures for only the Lower Bound, Most Likely Value, and the Upper Bound of activity durations. Statistical Dependence among the marginal distributions of activity durations is achieved by specifying a particular value, *Beta* – which in general drives the marginal uncertainty of activities. In the context of the procedure implemented in *Method 1*, it is important to note that the marginal activity duration uncertainty is not an effect realized *ab initio*, rather the marginal activity duration uncertainty is an output and realization of the procedure inherent in *Method 1*. The value of the shape parameter, *n,* which is used in *Method 1* is unique to each project activity.
- ***Method 2***: Application of *Method 2* is described in Chapter 4 as *Case Study 2*. The modified PERT variance (Herrerias-Velasco, Herrerias-Pleguezuelo, & van Dorp, 2011) and the classical PERT variance (Malcolm, Roseboom, Clark, & Fazar, 1959) are both applied in the technique described in this study as *Method 2*. In this method, even though we are specifying the average PERT variance over all the activities, the individual variances that we get for all the activities are very close to the modified PERT variance. Using a numerical procedure we specify two parameters at the same time, i.e. parameters *n* and β. Parameter, *n*, which

would be shared by all the activities; and parameter, β, which would be shared across the entire Bayesian Network.

In contrast with *Method 2*, the parameter *n* in Method 1 for all the activities differs by each activity. In the instance of Method 1, no additional information needs to be elicited because it follows from the values of *a, m, b* specified for the activities. We utilize a numerical procedure in Method 1 from which the value of *Beta* is specified from the Average Conditional Median approach.

For *Method 2*, we also utilize the Conditional Median, but in addition to that, we also calculate and use the average PERT variance – essentially, we are selecting two parameters (*n* and β) which are shared across all project activities. In this manner we compute for the desired average PERT variance over all the activities and also the average conditional median across the project network (as applied to the case where statistical dependence is incorporated among marginal distributions of activity durations).

In essence, the shape parameter *n* in Method 1 is unique to each project activity; however, in Method 2, the shape parameter n is shared among all the activities. One of the motivations for doing this is to examine how uncertainty in completion times both at the activity and project levels are affected if the parameter *n* is changed in such manner.

3.2 BN Model for *Method 1*

We describe below the elements and mechanics of the BN model for *Method 1*, i.e. the Common Quantile Level, the Mode Level, and the Duration Level.

3.2.1 Common Quantile Level

The application of quantile measures in risk analysis has become common practice in such areas as finance, Insurance, etc (Alexander & Sarabia, 2012). In the BN model, we use a non-informative and equi-probability distribution – the uniform distribution – as a monitoring variable that captures different possible values of programmatic risk which can always be transformed to a latent uniform random variable for the entire model, thus: $Y \sim U[0,1]$.

3.2.2 Mode Level

The Mode Level represents the Mode Relative Distance Uncertainty Model. To describe the mechanics we assume, for example, in a project activity network with 5 activities *A, B, C, D, and E* (shown in Figure 3-1), we represent X for each activity such that:

Activity X: Lower Bound = a_x
Upper Bound = b_x
Most Likely Value = m_x

Also, representing δ_x as the Relative Distance of the Most Likely Value of X, we have that:

$$\delta_x = \frac{m_x - a_x}{b_x - a_x} \tag{3-1}$$

A random variable Δ_x which represents the uncertainty in the relative distance of the Most Likely Value to the lower bound *a* given the width of the support [*a*, *b*]. We herein model that uncertainty given a programmatic risk common quantile measure, as a Two-Sided Power distribution (TSP) with parameters δ_x and n_x, thus: $(\Delta_x | Y) \sim TSP(\delta_x, n_x)$. The TSP distribution was proposed by van Dorp & Kotz (2002b) as an extension of the Triangular distribution (which has traditionally been used in problems involving risk and uncertainty, especially in the Program Evaluation and Review Technique, PERT). Some of the benefits proposed for the use of the TSP distribution include: 1) closed form derivation of their cumulative distribution function and its inverse, enabling its applicability to simulation problems in risk and uncertainty analysis; 2) computationally tractable maximum likelihood parameter estimation, etc (van Dorp & Kotz, 2002b). The parameterization of the density of the TSP distribution as shown in Equation 3-2 enables the interpretation of δ_x as the location parameter and n_x as the shape parameter, both of which determine the uncertainty in *X*.

Based on the description of the density function of the TSP distribution in van Dorp & Kotz (2002a), we have:

$$f_{\Delta_x}(\delta) \propto \begin{cases} \delta^{n_x - 1} & 0 < \delta < \delta_x \\ (1-\delta)^{n_x - 1} & \delta_x < \delta < 1 \end{cases} \tag{3-2}$$

$$Mode(\Delta_x) = \delta_x = \frac{m_x - a_x}{b_x - a_x} \tag{3-3}$$

$$V(\Delta_x) = \frac{n_x - 2(n_x - 1)\delta_x(1 - \delta_x)}{(n_x + 2)(n_x + 1)^2} \tag{3-4}$$

From the corresponding values of $\delta_A, \delta_B, \delta_C, \delta_D, \delta_E$; the value of each $n_x, x \in \{A, B, C, D, E\}$ can be computed from Equation 3-4 by solving from $V(\Delta_x) = s^2$ given that:

$$\bar{\delta} = \frac{1}{4} \times \sum_{i \in \{A,B,C,D,E)} \delta_i \tag{3-5}$$

$$s^2 = \frac{1}{4} \times \sum_{i \in \{A,B,C,D,E)} (\delta_i - \bar{\delta})^2 \tag{3-6}$$

3.2.3 Duration Level

The Duration Level represents the activity uncertainty given the relative distance of the Most Likely Value, Δ_x, here modeled as a Beta distribution (with parameters $a_x, b_x, \Delta_x, \beta$), thus: $(X \mid \Delta_x) \sim \text{Beta}(a_x, b_x, \Delta_x, \beta_x)$. Applying the parameters $a_x, b_x, \Delta_x, \beta$ to the density function of the Beta distribution as described in van Dorp & Mazzuchi (2004), we have:

$$f_x(\mathrm{x}) \propto (x-a_x)^{\beta\alpha_x-1}(b_x-x)^{\beta_x(1-\alpha_x)-1} \tag{3-7}$$

$$V[X] = \frac{\alpha_x(1-\alpha_x)}{\beta_x+1} \times (b_x-a_x)^2 \tag{3-8}$$

$$Mode[X] = \frac{\beta_x\alpha_x-1}{\beta-2} \times (b_x-a_x) + a_x \tag{3-9}$$

Parameterizing Equation 3-9 with $\frac{\beta_x\alpha_x-1}{\beta-2} = \Delta_x$, we have that:

$$Mode[X] = \Delta_X \times (b_x-a_x) + a_x \tag{3-10}$$

Solving for α_x from $\frac{\beta_x\alpha_x-1}{\beta-2} = \Delta_x$ implies that:

$$\alpha_x = \frac{\Delta_x(\beta-2)+1}{\beta_x} \tag{3-11}$$

The reparameterization described from (3-7) to (3-11) is indicative that α and β are location and uncertainty parameters respectively, which also describe the uncertainty in X. From the description above, it follows that the only parameters that need to be specified is β_x. We shall set $\beta_x = \beta$ such that β represents a measure of common variance amongst the activities. It is important to point out that only the parameter β is specified for the degree of statistical dependence among the marginal distribution of activities. More so, the marginal activity variance derived in *Method 1* is an output of the procedure implemented herein as *Method 1*, and not a condition realized *ab initio*. In sections 3.3 and 3.4, we describe the indirect elicitation procedure for β.

3.3 AgenaRisk® representation of the BN model for *Method 1*

Shown below in Figure 3-3 is a representation of the BN model for *Method 1* (Figure 3-2) as implemented in *AgenaRisk®* for the 5-activity (*A, B, C, D, and E*) project network shown in Figure 3-1. Essentially, Figure 3-3 is a BN representation of the BN model shown in Figure 3-2.

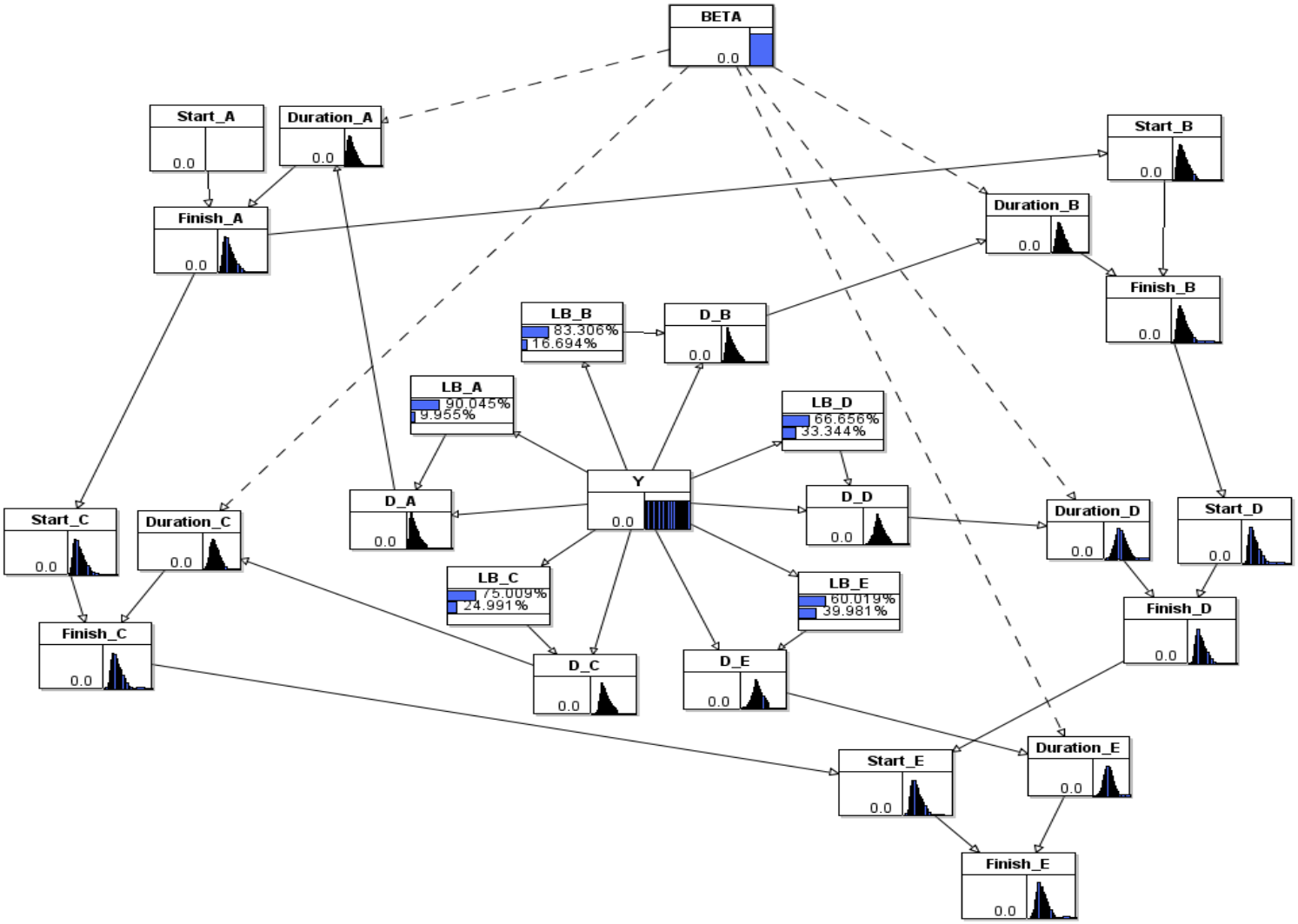


Figure 3-3: BN model for Method 1 implemented in AgenaRisk for the example project for Case Study 1

The BN representation (Figure 3-3) of the example project network (Figure 3-1) inherently follows the Activity-on-Node (AoN) scheduling technique. It also consists of the 3 levels (i.e. the Common Quantile Level; the Mode Level; and the Duration Level) discussed earlier in Section 3.1.

As with the BN model, the representation in Figure 3-3 also consists of 3 levels: a) The **Common Quantile Level** consists of only the *Y* node; b) The **Mode Level** consists of 10 nodes: D_A and LB_A for activity A, D_B and LB_B for activity B, D_C and LB_C for activity C, D_D and LB_D for activity D, and D_E and LB_E for activity E; c) For each project activity: A, B, C, D, E the **Duration Level** consists of the *Start* node, the *Finish* node, and the *Duration* node. A description of the nodes is shown below:

- Start node: In the BN representation of Figure 3-3, we effectively model each project activity (of the example project network of Figure 3-1) with 3 nodes: the *Start* node, the *Finish* node, and the *Duration* node. The *Start* node effectively represents the *Early Start* (ES) of a project activity – modeled in AgenaRisk® as a continuous node using *Arithmetic Expression*.
- Finish node: The *Finish* node effectively represents the *Early Finish* (EF) of a project activity – modeled in AgenaRisk® as a continuous node using *Arithmetic* Expression.
- Duration node: The *Duration* node represents the duration of a project activity – modeled in AgenaRisk® as a Beta distribution with lower and upper bounds which effectively represent duration parameters *a* and *b* respectively. With

reference to the reparameterization of the Beta distribution described in (van Dorp & Mazzuchi, 2004) using $a = \alpha\beta; b = (1-\alpha)\beta; \alpha = \frac{a}{a+b}$, we determine in AgenaRisk® (using *Expressions*) the estimates of the parameters α and β of the Beta distribution – (please refer to Equation 3-11).

- *Y* node: This effectively represents the common quantile, modeled as a uniform distribution with bound [0, 1].
- Mode Level nodes: The Mode Level nodes are represented by ***D_X*** and ***LB_X*** (i.e. in the example project network, "*X*" represents each project activity: A, B, C, D, and E). Essentially, using both the *D_X* and *LB_X* we have implemented the equation: $F^{-1}(Y) = \begin{cases} L + \sqrt[n]{y(U-L)(M-L)^{n-1}}, & 0 \le y \le \frac{M-L}{U-L} \\ U - \sqrt[n]{(1-y)(U-L)(U-M)^{n-1}}, & \frac{M-L}{U-L} \le \delta \le 1 \end{cases}$ which represents an alternative to *B*eta distribution (Kotz & van Dorp, 2004).
- *Beta*: The node indicated as *Beta* is the factor that drives statistical dependence across the durations of activities of the project network.

3.4 Activity Statistical Dependence

We describe the mechanism for specifying degree of dependence by ensuring that the following conditions are satisfied:

- Low Degree of Dependence: In this case, the degree of dependence decreases as:

$$\beta = \frac{1}{\alpha_X} \Rightarrow X \sim U[0,1] \Rightarrow Corr(X,Y) = 0, Y \in \{A,B,C,D,E\}, Y \neq X$$

- High Degree of Dependence: In this case, the degree of dependence increases: $\beta \to \infty$. Thus:

  $\beta \to \infty \Rightarrow X = \Delta_x, X \in \{A, B, C, D, E) \Rightarrow$

  $$Corr(X, Y) = 1, X \in \{A, B, C, D, E), Y \in \{A, B, C, D, E)$$

Based on the conditions mentioned above, the overall interpretation, therefore, is that intermediate values of *Beta* lead to intermediate degree of dependence.

3.5 Eliciting Degree of Dependence

We propose the use of conditional median as a parametric solution for location problems involving conditional distribution functions. Chaloner & Duncan (1987; 1983) mention the use of quantiles, for example the median, for expert elicitation purposes, and van Dorp & Mazzuchi (2004) describe the efficient use of quantiles or conditional median and its suitability in application areas such as in Bayesian Analysis. With respect to the BN model developed in this study, the conditional median is used herein as a statistic for encoding statistical dependence among project activities. It is important to note that in order to avoid anchoring bias, the elicitation of the conditional median is performed in a randomized manner.

3.5.1 Eliciting Conditional Medians

As discussed earlier, the impact of programmatic risk factors (for example, requirement changes) can be experienced at any stage of a project life cycle. Suppose at the event of requirement changes in a project, we herein model the supposed impact of these

events on activity durations by utilizing conditional median to determine the level of probabilistic relationship across activities in a project network – this is implied both for *Method 1* and *Method 2*. Van Dorp & Mazzuchi (2004) mention that elicitation procedures for determining conditional median from experts are more practical and require less cognitive processing, and therefore introduces less random noise to the analysis process as compared to other statistical measures. Also, the procedure for the elicitation of conditional median discussed in this study (both for *Method 1* and *Method 2*) is equivalent to the indirect elicitation of Blomquist's β discussed in Kotz & van Dorp (2010). The Blomquist's β is one of the lesser known measures of association reminiscent of Kendall's tau and Spearman's Rank Correlation.

We refer to our example 5-activity project network, here reproduced as an Activity-on-Arrow project network:

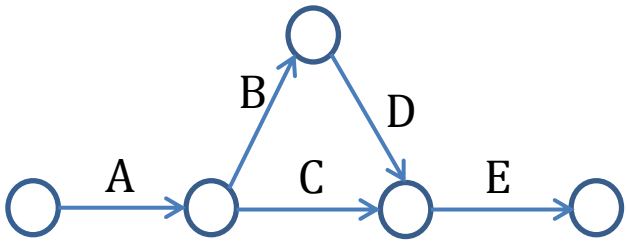


Figure 3-4: AoA representation of example project network

With reference to Figure 3-4, our elicitation procedure or protocol follows from the question: *Suppose activity A has finished above its median value x, what is the probability that activity B finishes above its median value y?*

Elicited responses are categorized thus:

- If answer is "= 0.50," knowledge of activity A does not influence B

- If answer is in (0.50, 1.0), knowledge of activity A partially influences B
- If answer is " = 1.00," knowledge of activity A completely influences B

It is important to point out that the elicitation procedure follows the path dependence structure of the project network. Also, for all activities of the project network, the elicitation procedure follows the same manner as expressed above, thus for Figure 3-4 we elicit:

1) $\Pr\{B > \text{Med}(B) \mid A > \text{Med}(A)\}$
2) $\Pr\{C > \text{Med}(C) \mid A > \text{Med}(A)\}$
3) $\Pr\{D > \text{Med}(D) \mid B > \text{Med}(B)\}$
4) $\Pr\{E > \text{Med}(E) \mid C > \text{Med}(C)\}$
5) $\Pr\{E > \text{Med}(E) \mid D > \text{Med}(D)\}$

3.5.2 Conditional Median Matrix

We determine the prior conditional median matrix by selecting a value for developing a joint sample of 1000 replications representing durations of project activity A, B, C, D, and E (van Dorp J. R., 2013). Table 3-1 shows a general representation or format of the conditional median matrix thus generated.

Table 3-1: Conditional Median Matrix

	A	B	C	D	E
A	1	Med(B|A)	Med(C|A)	Med(D|A)	Med(E|A)
B	Med(A|B)	1	Med(C|B)	Med(D|B)	Med(E|B)
C	Med(A|C)	Med(B|C)	1	Med(D|C)	Med(E|C)
D	Med(A|D)	Med(B|D)	Med(C|D)	1	Med(E|D)
E	Med(A|E)	Med(B|E)	Med(C|E)	Med(D|E)	1

Notation: Pr {Y > Med (Y) | X > Med (X)} = Med(Y|X)

With respect to Table 3-1 shown above, the elements shown in red color reflect the activity-to-activity conditional median based on the path-dependent structure of the example project network of Figure 3-1. It can be seen from equation 3-12 that the Conditional Median Matrix is symmetric, thus:

$$\Pr\{X > \text{Med}(X) \mid Y > \text{Med}(Y)\} = \frac{\Pr\{Y > \text{Med}(Y) \mid X > \text{Med}(X)\}\Pr\{X > \text{Med}(X)\}}{\Pr\{Y > \text{Med}(Y)\}}$$

$$= \frac{\Pr\{Y > \text{Med}(Y) \mid X > \text{Med}(X)\} * 0.5}{0.5}$$

$$= \Pr\{Y > \text{Med}(Y) \mid X > \text{Med}(X)\} \qquad (3\text{-}12)$$

3.5.3 Average Conditional Median

We evaluate the Average Conditional Median from the conditional median matrix described in Table 3-1 by implementing the formula:

$$\overline{Med(\cdot|\cdot)} = \frac{\sum_{\substack{Y\in\{A,B,C,D,E\} \\ Y\neq X}} \sum_{X\in\{A,B,C,D,E\}} Med(Y \mid X)}{n(n-1)} \tag{3-13}$$

In equation 3-13, "*n*" represents the number of activities in a project network. In selecting the remaining parameters of *Beta*, β, we utilize the following algorithmic steps:

1. Select initial (high) value of *Beta*
2. Use prior sampling to evaluate conditional median matrix
3. Evaluate average conditional median
4. If average conditional median is "close" to elicited average conditional median then

   Stop

   Else

   Set β = β/2 and Goto Step 2

It is important to note that the notion of averaging the conditional median as discussed herein is premised on the understanding that we want to hedge against or compensate for the random fluctuations (or noise) that might arise resulting from elicitation protocol utilized in determining the activity-to-activity conditional median elicited from the experts. This viewpoint is also reflected in the concept of average conditional median discussed in Section 4.5.

3.6 Case Study 1

Case Study 1 describes the implementation of the BN model for *Method 1* using dependence specification procedure that involves mode relative distance uncertainty. A project network involving 5 activities shown previously in Figure 3-1 is used in the analysis, and discussed below is the input data used in the analysis of the case study.

Input Data: Using an Excel Spreadsheet Dependent Sampler (van Dorp J. R., 2013), we develop model input data that controls for dependence parameters of the BN model implemented in AgenaRisk® for varying levels of *Beta*, that is, β = 40, 20, and 10. For each level of *Beta*, we determine the degree of dependence by also conducting analysis of the corresponding correlation matrix and conditional median matrix. The tables shown below describe the results thereof:

Table 3-2: Case Study 1 input data and model parameters for β *= 40. (For prior sampling repeat Step 2 and sample from Beta's)*

STEP 1

	Elicited Activity Parameters			Δ Parameters		
X	a_x	m_x	b_x	$Mode[\Delta_x]$	$Var(\Delta_x)$	n_x
A	4	5	14	0.1000	0.0147	5.4395
B	3	4	9	0.1667	0.0147	5.1144
C	8	10	16	0.2500	0.0147	4.6175
D	1	2	4	0.3333	0.0147	4.4224
E	3	5	8	0.4000	0.0147	4.3192

STEP 2

	Common Quantile		BETA Parameters for Sampling		
X	Y	Δ Sample	Mode[X]	α_x	β
A	0.7712	0.3003	0.3003	0.3103	40
B	0.7712	0.3528	0.3528	0.3601	40
C	0.7712	0.4200	0.4200	0.4240	40
D	0.7712	0.4765	0.4765	0.4777	40
E	0.7712	0.5200	0.5200	0.5190	40

Table 3-3: Case Study 1 input data and model parameters for β = 40 showing high degree of dependence and Average Conditional Median, $\overline{Med}(\cdot|\cdot) \approx 0.746$

Traditional Correlation Matrix (β = 40)

	A	B	C	D	E
A	1.00	0.78	0.76	0.74	0.73
B	0.78	1.00	0.74	0.74	0.73
C	0.76	0.74	1.00	0.74	0.72
D	0.74	0.74	0.74	1.00	0.71
E	0.73	0.73	0.72	0.71	1.00

Conditional Median Matrix (β = 40)

	A	B	C	D	E
A	1.00	0.77	0.77	0.75	0.75
B	0.77	1.00	0.74	0.73	0.75
C	0.77	0.74	1.00	0.74	0.73
D	0.75	0.73	0.74	1.00	0.72
E	0.75	0.75	0.73	0.72	1.00

Table 3-4: Case Study 1 input data and model parameters for β = 20. (For prior sampling repeat Step 2 and sample from Beta's)

STEP 1

	Elicited Activity Parameters			Δ Parameters		
X	a_x	m_x	b_x	$\mathrm{Mode}[\Delta_x]$	$\mathrm{Var}(\Delta_x)$	n_x
A	4	5	14	0.1000	0.0147	5.4395
B	3	4	9	0.1667	0.0147	5.1144
C	8	10	16	0.2500	0.0147	4.6175
D	1	2	4	0.3333	0.0147	4.4224
E	3	5	8	0.4000	0.0147	4.3192

STEP 2

	Common Quantile		BETA Parameters for Sampling		
X	Y	Δ Sample	Mode[X]	α_x	β
A	0.7712	0.3003	0.3003	0.3203	20
B	0.7712	0.3528	0.3528	0.3675	20
C	0.7712	0.4200	0.4200	0.4280	20
D	0.7712	0.4765	0.4765	0.4789	20
E	0.7712	0.5200	0.5200	0.5180	20

Table 3-5: Case Study 1 input data and model parameters for β = 20 showing mild degree of dependence and Average Conditional Median, $\overline{Med}(\cdot|\cdot) \approx 0.673$

Traditional Correlation Matrix (β=20)

	A	B	C	D	E
A	1.00	0.58	0.60	0.56	0.56
B	0.58	1.00	0.59	0.53	0.56
C	0.60	0.59	1.00	0.57	0.56
D	0.56	0.53	0.57	1.00	0.52
E	0.56	0.56	0.56	0.52	1.00

Conditional Median Matrix (β=20)

	A	B	C	D	E
A	1.00	0.68	0.69	0.68	0.68
B	0.68	1.00	0.69	0.64	0.67
C	0.69	0.69	1.00	0.67	0.68
D	0.68	0.64	0.67	1.00	0.65
E	0.68	0.67	0.68	0.65	1.00

Table 3-6: Case Study 1 input data and model parameters for β = 10. (For prior sampling repeat Step 2 and sample from Beta's).

STEP 1

	Elicited Activity Parameters			Δ Parameters		
X	a_x	m_x	b_x	$Mode[\Delta_x]$	$Var(\Delta_x)$	n_x
A	4	5	14	0.1000	0.0147	5.4395
B	3	4	9	0.1667	0.0147	5.1144
C	8	10	16	0.2500	0.0147	4.6175
D	1	2	4	0.3333	0.0147	4.4224
E	3	5	8	0.4000	0.0147	4.3192

STEP 2

	Common Quantile		BETA Parameters for Sampling		
X	Y	Δ Sample	Mode[X]	α_x	β
A	0.7712	0.3003	0.3003	0.3403	10
B	0.7712	0.3528	0.3528	0.3822	10
C	0.7712	0.4200	0.4200	0.4360	10
D	0.7712	0.4765	0.4765	0.4812	10
E	0.7712	0.5200	0.5200	0.5160	10

Table 3-7: Case Study 1 input data and model parameters for β = 10 showing low degree of dependence and Average Conditional Median, $\overline{Med}(\cdot|\cdot) \approx 0.604$

Traditional Correlation Matrix (β=10)

	A	B	C	D	E
A	1.00	0.37	0.35	0.40	0.32
B	0.37	1.00	0.36	0.33	0.32
C	0.35	0.36	1.00	0.32	0.34
D	0.40	0.33	0.32	1.00	0.32
E	0.32	0.32	0.34	0.32	1.00

Conditional Median Matrix (β=10)

	A	B	C	D	E
A	1.00	0.59	0.60	0.61	0.60
B	0.59	1.00	0.59	0.58	0.59
C	0.60	0.59	1.00	0.61	0.62
D	0.61	0.58	0.61	1.00	0.63
E	0.60	0.59	0.62	0.63	1.00

3.7 Prior Analysis

We perform prior analysis of the 5-activity project network of Figure 3-4 with input data for β = 10. Shown below in Table 3-8 are the percentile values of durations of the project activities – A, B, C, D, and E. In computing for the percentile values; firstly, we set the value for β in the BN model of Figure 3-2 (which is implemented in AgenaRisk®) and export (as CSV file) the activity duration marginals and the project completion distribution; and secondly, using Least Squares method implemented in an Excel

spreadsheet for the project activity duration marginals (van Dorp & Nduka, 2013), we deduce the parameters of the Beta fitted marginal distributions of activity durations and their respective percentiles as shown in Table 3-8.

Table 3-8: Percentile values for project activity durations

	A	B	C	D	E
LB	4	3	8	1	3
UB	14	9	16	4	8
alpha	2.56043	3.48069	4.37290	5.39939	6.05032
beta	8.67472	9.06449	8.51172	8.26937	7.76967
5th percentile	4.61771	3.59644	9.16249	1.58153	4.15315
50th percentile	6.11391	4.59199	10.64680	2.16927	5.17355
95th percentile	8.50814	5.98313	12.50266	2.84284	6.27781

Using the 5^{th} percentile instantiation values from Table 3-8 we develop the risk profiles or the Project Completion Time Cumulative Distribution Function (CDF) shown in Figure 3-5 for the 5-activity project network. Representation of the risk profiles for both the independence and dependence scenarios are shown in Figure 3-5. It can be interpreted upon observation of the risk profiles shown in Figure 3-5 for both the Dependence and Independence scenarios that there is greater spread in the distribution of the Dependence scenario, which is indicative of greater uncertainty when compared to the distribution of the Independence scenario. This means that without modeling for dependence as we have implemented using the BN model for *Method 1*, there is the tendency to underestimate the duration uncertainty of projects.

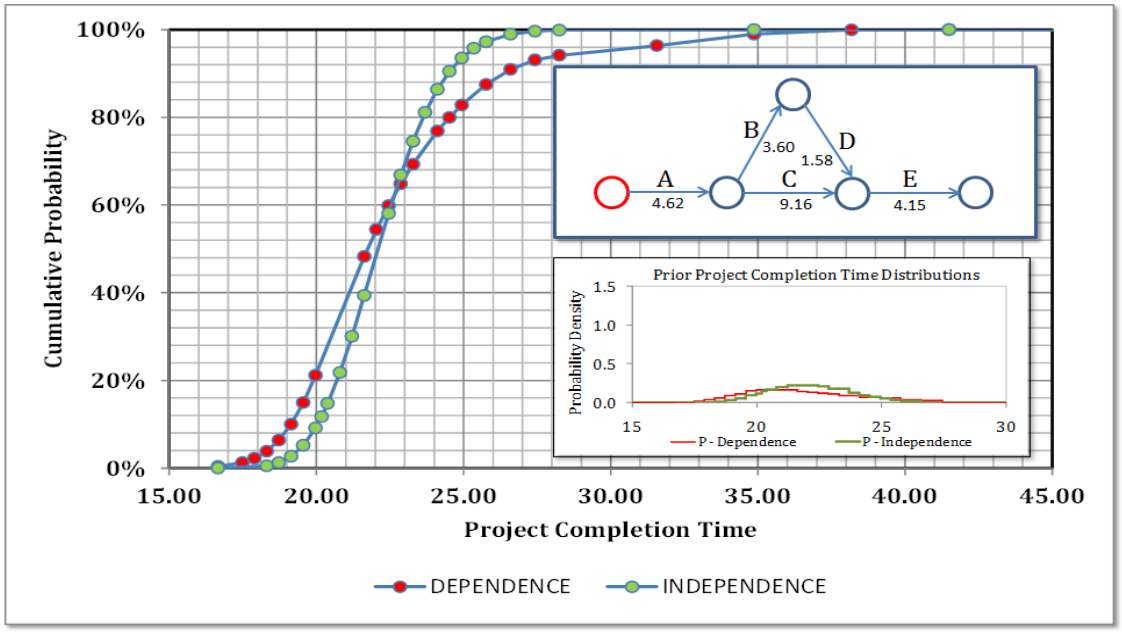


Figure 3-5: Project Completion Time CDF without Instantiation of Activity Duration Nodes

It is important to note that in the context of the discussion contained in this study, the Independence and Dependence scenarios as mentioned in Figure 3-5 (and subsequently in this document) mean:

- Independence scenario: a representation of a project network where statistical independence among the marginal distributions of activity durations is implied. In this instance, the elements of the BN model (i.e. Common Quantile Level, the Mode Level, and *Beta*) are not incorporated.
- Dependence scenario: a representation of a project network with statistical dependence implemented among the marginal distributions of activity durations. In this instance, all the elements and parameters of the BN model are incorporated in the representation and modeling.

It is important to point out some semantic nuances here: the classical PERT project network, by design, encodes some kind of precedence relationship, and therefore, the dependence therein is assumed as functional (no statistical dependence is implied). However, the dependence implied in the BN models (i.e. Method 1 and Method 2) developed in this study is as a result of using the Bayesian paradigm to incorporate the statistical dependence among the marginal distributions of activity durations – the dependence in this case, to speak in statistical terms, is as a result of association.

We illustrate below in Figure 3-6, the AgenaRisk® representation of the 5-activity project network for both the Independence and Dependence scenarios.

• Project network (Independence scenario)

• Project network (Dependence scenario)

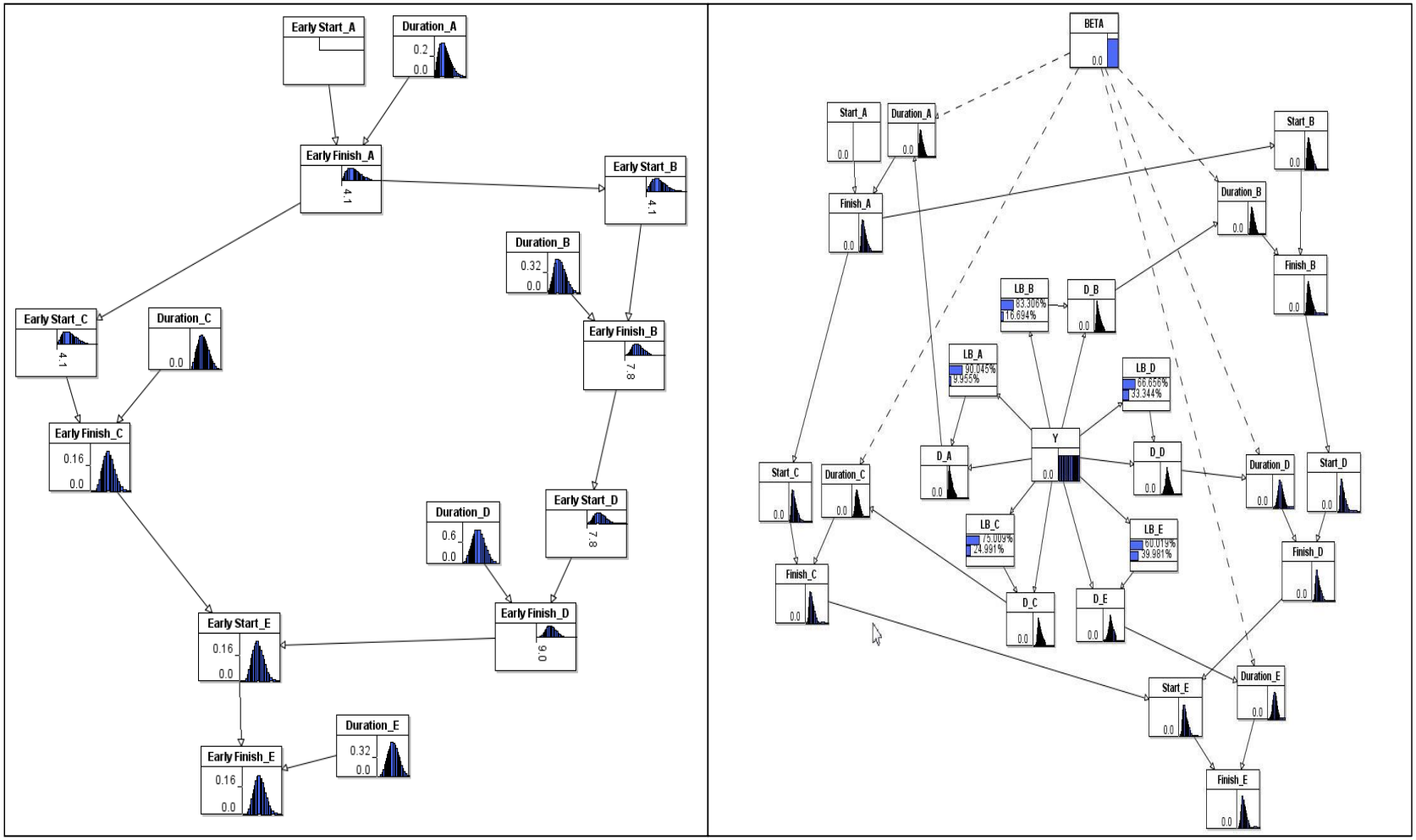


Figure 3-6: AgenaRisk® representation of the BN model for Method 1 (showing comparison with implementation for the Independence scenario)

3.8 Posterior Analysis

In performing posterior analysis both for the dependent and independent scenarios using Method 1, we instantiate at the Duration nodes of the BN model (both for the dependent and independent scenarios) under the assumption that "*if we had known at the start of the project the completion time of certain activities, what then would be the remaining completion time uncertainty as measured from the start of the project*." It is important to point out that the aforementioned assumption which is used in conducting the posterior analysis both for Method 1 (and also for Method 2) is different from a situation where the posterior analysis is conducted with the assumption that: "*At the time of the completion of an activity what is the remaining completion time uncertainty of the project from the completion time of that activity onward*."

We perform posterior analysis by instantiating the duration nodes with the 5^{th}, 50^{th}, and 95^{th} percentile values from Table 3-8.

3.8.1 Posterior Analysis of the Monitoring Variable (Common Quantile Level)

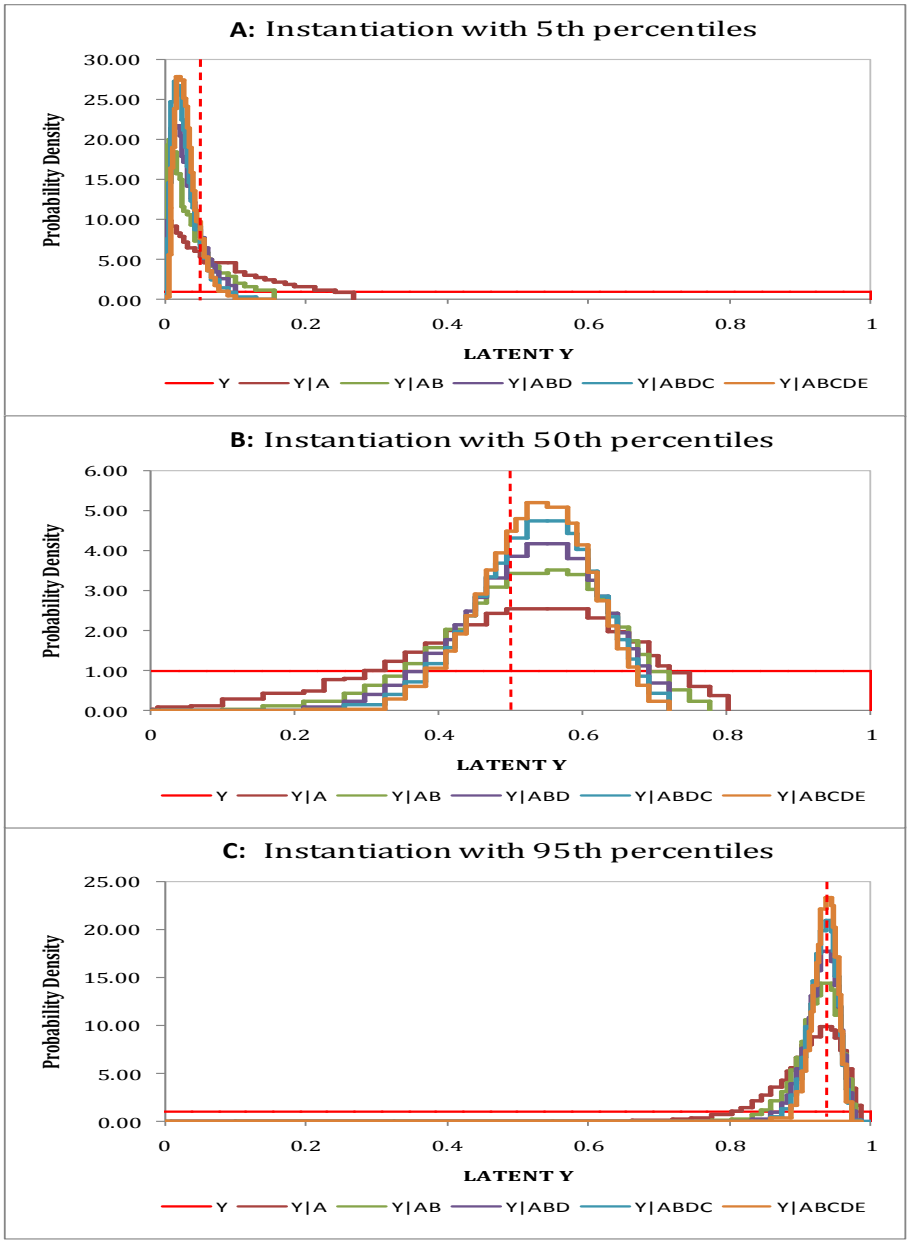


Figure 3-7: Distribution of the Common Quantile Level (Monitoring Variable) after instantiation with: A) 5th, B) 50th, and C) 95th percentiles of activity duration values

Figure 3-7 shown above represents a combination of distributions of the Common Quantile Level representing programmatic risk in the project before and after instantiation with 5th, 50th, and 95th percentiles of activity duration values. The

distributions show that the monitoring variable, Y, has specified uniform distribution *a priori* across the activities i.e. $Y \sim U[0,1] \Rightarrow$ *A priori* A, B, C, D, and E have specified marginal distribution. However, after instantiation with different percentile values, we have that the marginal distributions of the activities differ remarkably i.e. *A posteriori* $Y \neq U[0,1] \Rightarrow$ *A posteriori* A, B, C, D, and E differ from specified marginal distributions.

It can be discerned from the risk profiles of Figure 3-7 that, *a posteriori*, the distributions converge to a value approximately equal to the percentile value of instantiation. For example, after instantiation with 95th percentile values, the distributions of $(Y|A)$, $(Y|AB)$, $(Y|ABD)$, $(Y|ABDC)$, $(Y|ABDCE)$ converge at a value approximately equal to 95% (or 0.95). This, in our best opinion, provides some level of validity to the analysis performed. From Figure 3-7, the intuition from the values at which the distributions of the Monitoring Variable converge is that it reflects the probability of attaining a project completion time after all project activities in the model have been instantiated.

3.8.2 Posterior Analysis of the Activity Duration Level

In Figure 3-8 we further demonstrate the effect of statistical dependence by examining the marginal distributions of activities as activities of the project network are instantiated.

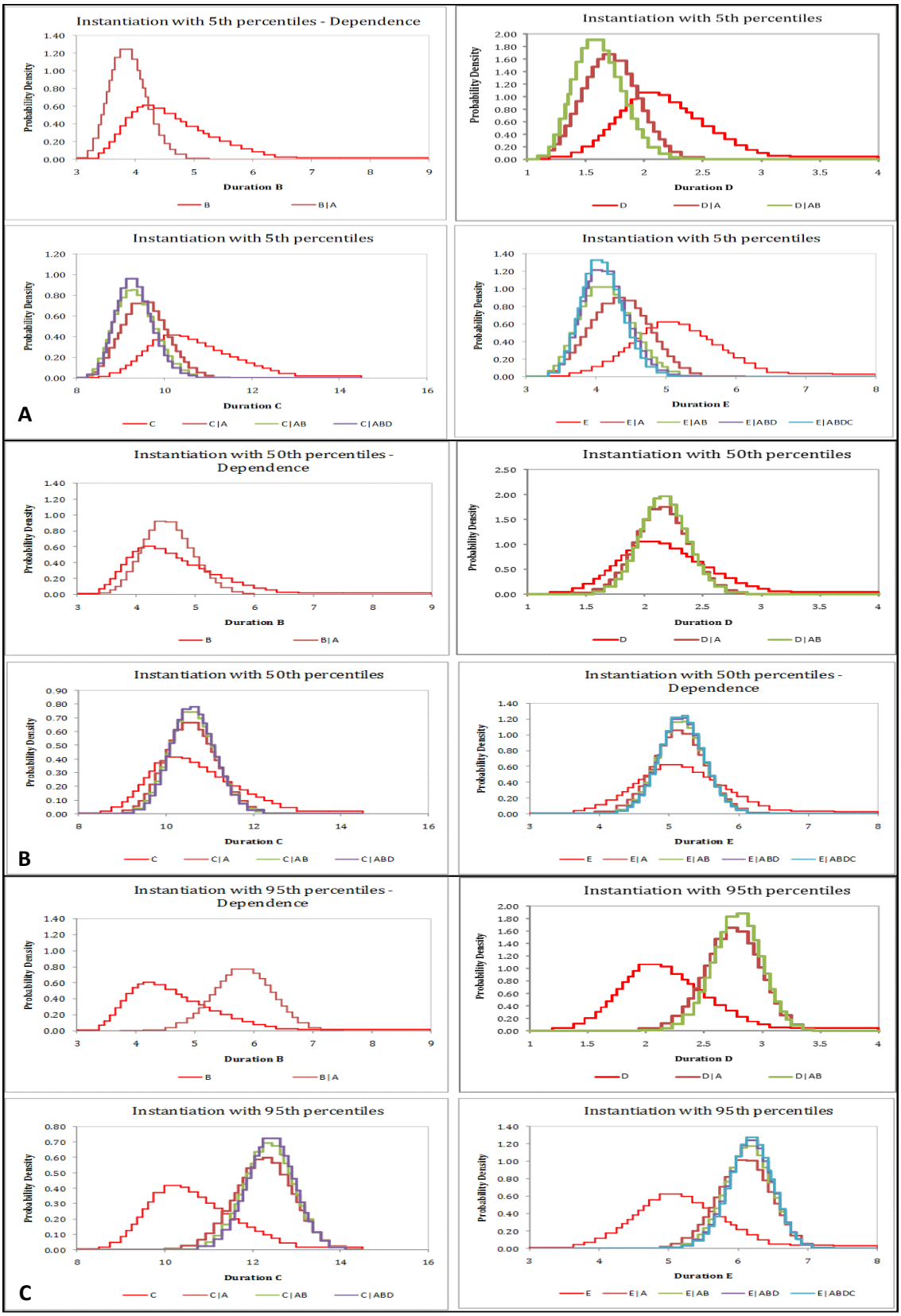


Figure 3-8: Distributions of activity duration (given the instantiations of activities across the project) for: A) 5th, B) 50th, and C) 95th percentile values

It is discernable from Figure 3-8 that there is progressive reduction in the variance of marginal distributions of project activities as activities are instantiated. This effect causes learning or provides some insight into the uncertainty of activity durations as activities across the project network are instantiated.

3.8.3 Posterior Analysis of Project Completion Times

In Figures 3-9 and 3-10 shown below we show a graphical description of results of analysis of effect of statistical dependence on project completion times by comparing the risk profiles for the 95^{th} percentile values of activity duration, both for the dependence and independence scenarios.

- Analysis of risk profiles of project completion times: By comparing the CDFs of Figure 3-9 (after instantiating at *Activities A, B,* and *D*), it could be discerned that the CDFs or risk profiles for the independence scenario remains constant (as activities are instantiated across the project from *Activities A, B, to D*), as opposed to the CDFs for the dependence scenario which is dynamic (as activities are instantiated across the project from *Activities A, B, to D*). The dynamism of the CDFs or risk profiles of the dependence scenario (across the project from *Activities A, B, to D*) underpins the effect of incorporating statistical dependence in the modeling process.

  However, as *Activity C* (whose CDF, hitherto, had remained constant) is instantiated, we notice a sharp movement of the CDF of the independence scenario (refer to Figure 3-10). This sharp lateral movement of the CDF of the independence scenario

as *Activity C* is instantiated is attributed to the high probability of *Activity C* being critical or in the longest path of the project network.

Finally, as *Activity E* is instantiated, the CDF for both the dependence and independence scenarios converge (as shown in Figure 3-10). The interpretation of the convergence of the risk profiles of both the dependence and independence scenarios as *Activity E* is instantiated indicates that uncertainty has been removed from the analysis and full learning or knowledge of project completion times has been accomplished up to that particular activity of instantiation.

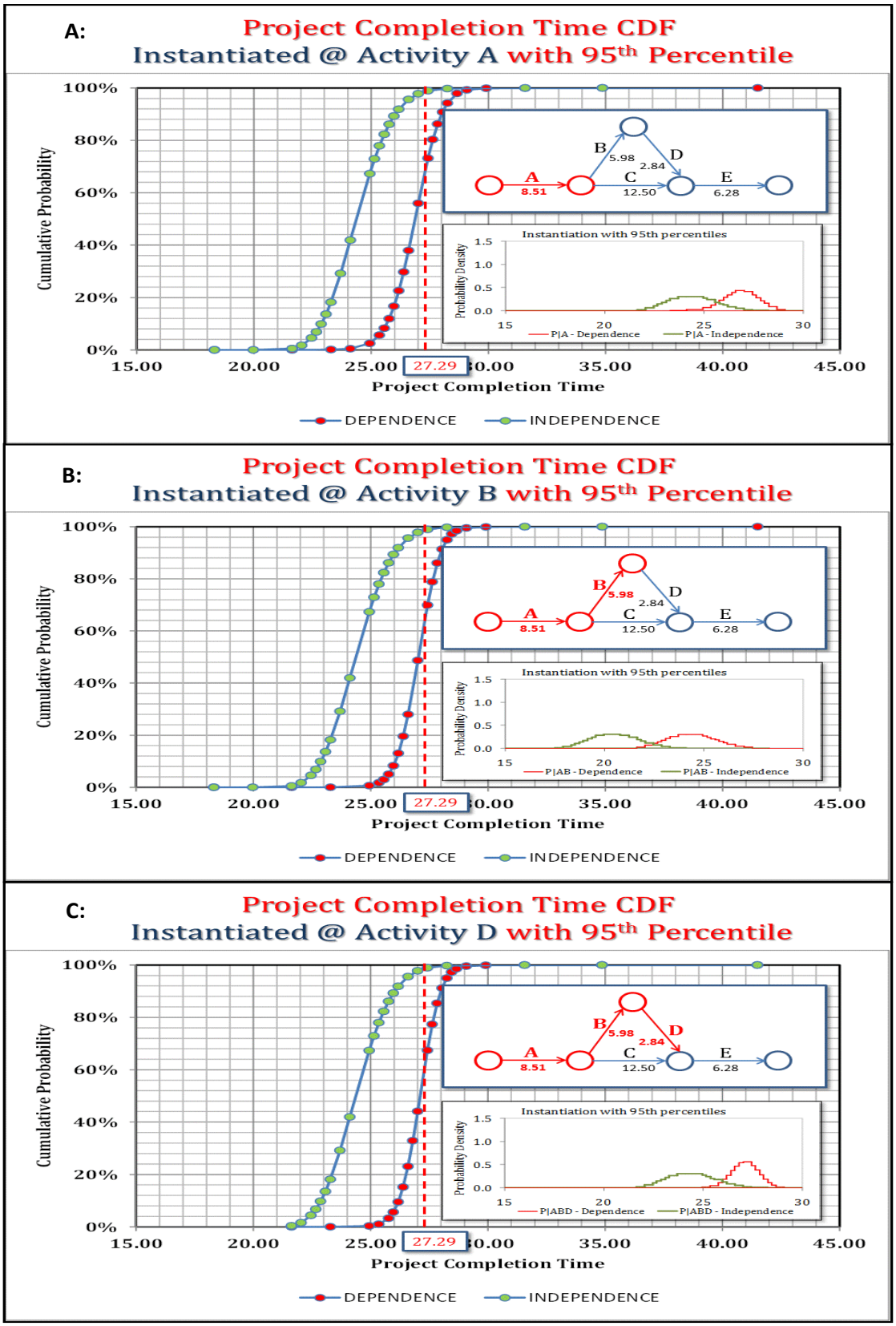


Figure 3-9: Project completion time CDF (instantiated with 95th percentile values) for both the independent and dependent scenarios

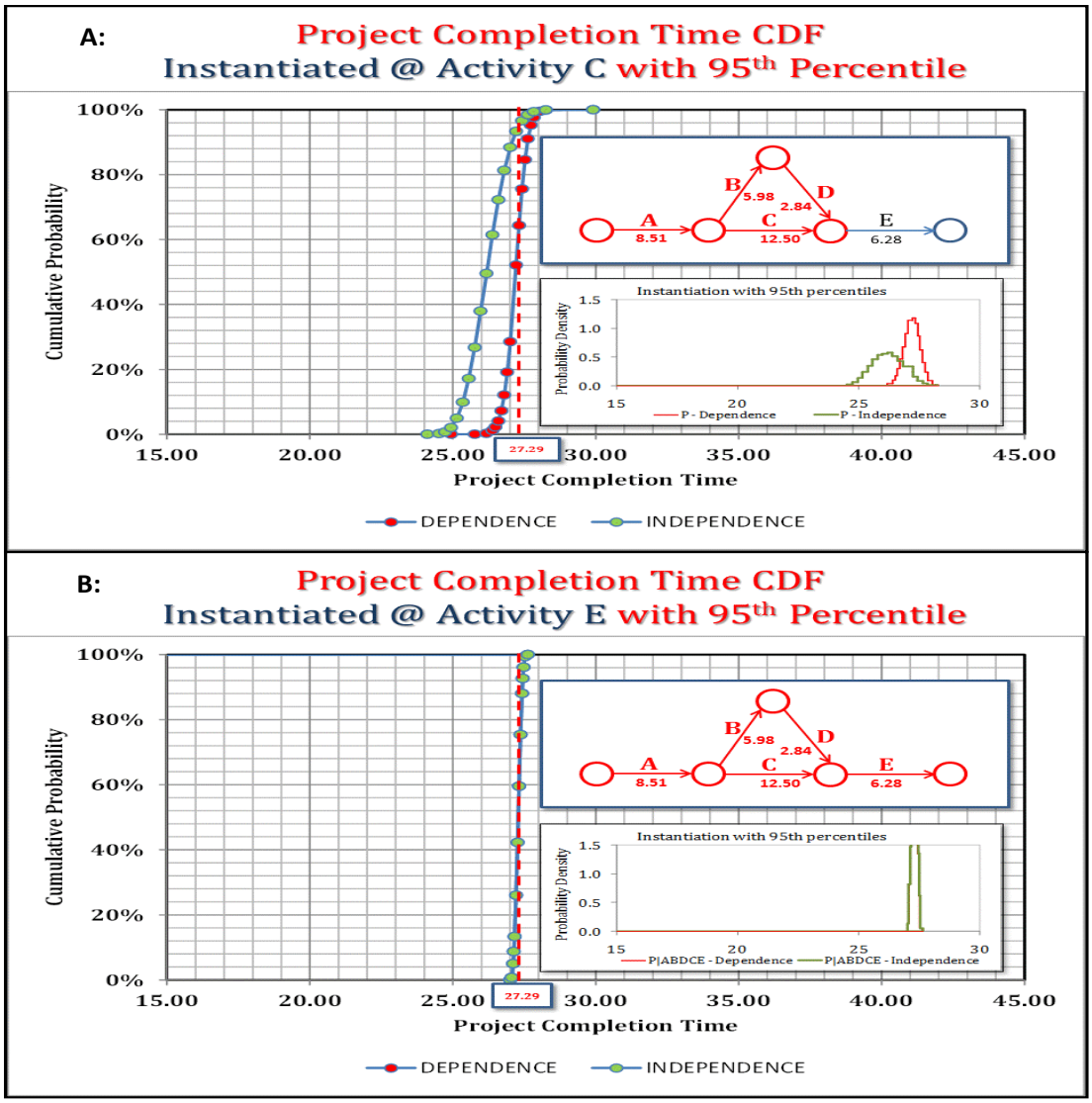


Figure 3-10: Project completion time CDF (instantiated with 95th percentile values) for both the independent and dependent scenarios

In Figure 3-11 shown below, we broaden the analysis of results by also describing and combining results for 5th, 50th, and 95th percentiles values of instantiation of activity duration. Analysis of Figure 3-11 indicates similar pattern of behavior of risk profiles (as described earlier for the 95th percentile value of instantiation) of the dependence and independence scenarios for activity duration instantiations using 5th, 50th, and 95th percentile values.

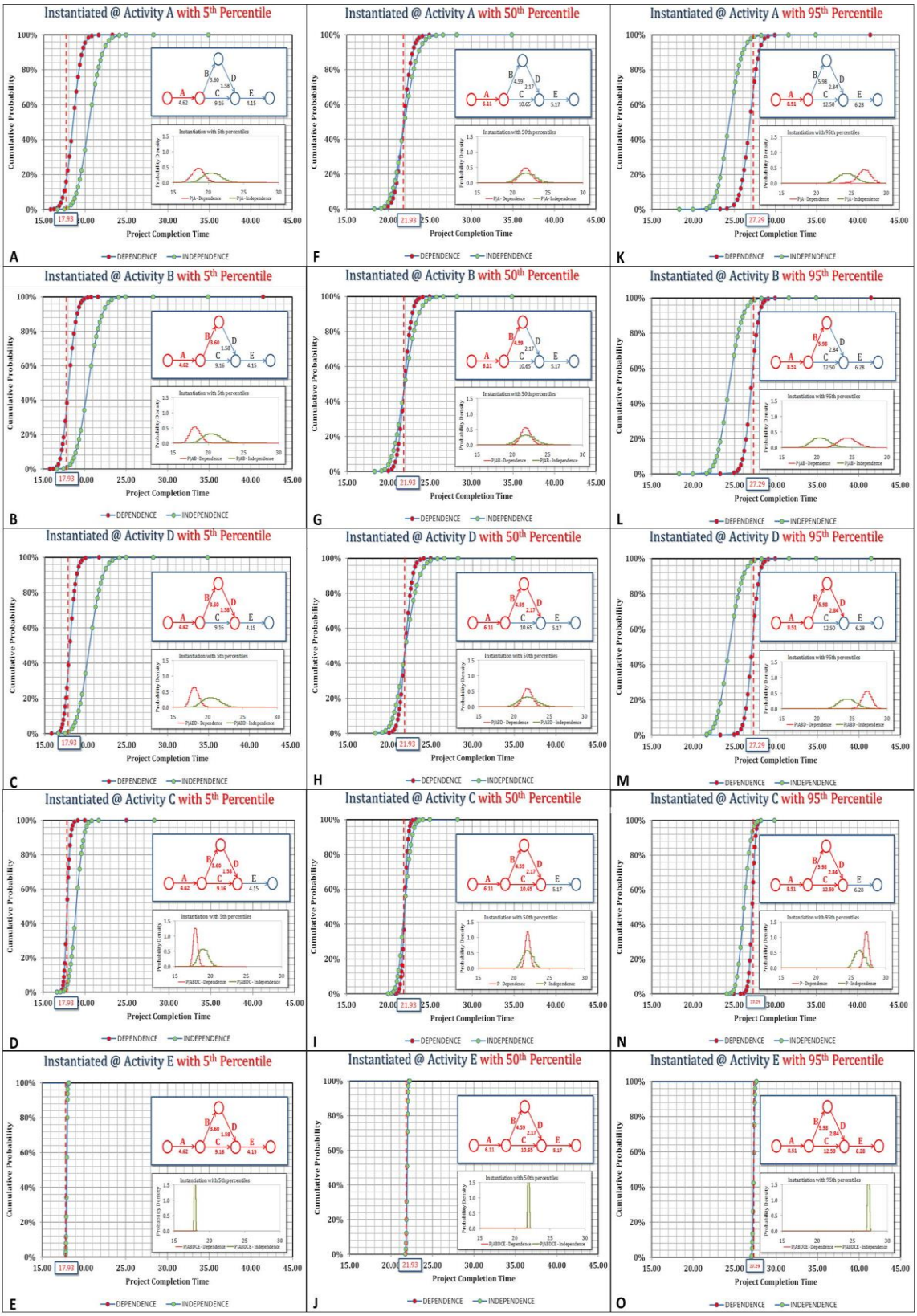


Figure 3-11: Project completion time CDF (with instantiation values as 5th, 50th, and 95th percentiles) for both the independent and dependent scenarios

3.8.4 Posterior Analysis of Project Completion Times (given evidence at activity duration nodes)

Using the example 5-activity project network, we demonstrate the effect of statistical dependence on project duration uncertainty (using 5th, 50th, and 95th percentiles of activity duration estimates) by describing the distributions of project completion times as project activities are progressively instantiated across the project network from project start to project finish [i.e. $(P|A)$, $(P|AB)$, $(P|ABD)$, $(P|ABDC)$, $(P|ABDCE)$].

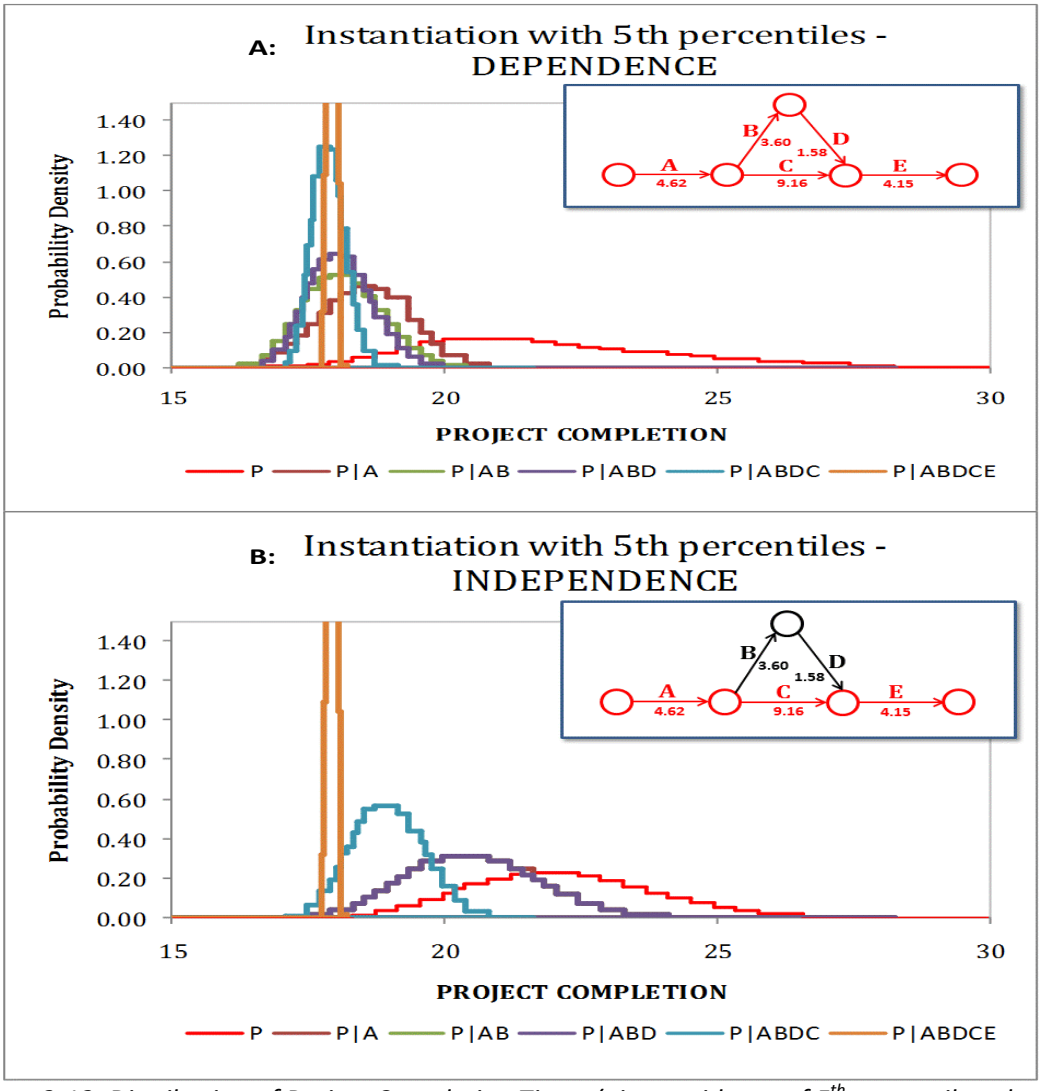


Figure 3-12: Distribution of Project Completion Times (given evidence of 5th percentile values of activity duration estimate across the project network)

As project activities are instantiated with different percentile values of duration estimates across the project network, we describe the analysis graphically in Figure 3-12A that activity duration uncertainty is clearly evident in the case of the dependence scenario than in the independence scenario. For example, in Figure 3-12A, the marginal distributions of durations for all project activities are obvious in the dependence scenario than in the independence scenario, which underscores the point of view that learning or insight about activity duration uncertainty is enhanced in the dependence scenario than in the independence scenario. Similar insight can be inferred when the project network is instantiated with 50^{th} and 95^{th} percentile values of activity duration estimates, as shown in Figures 3-13 and 3-14.

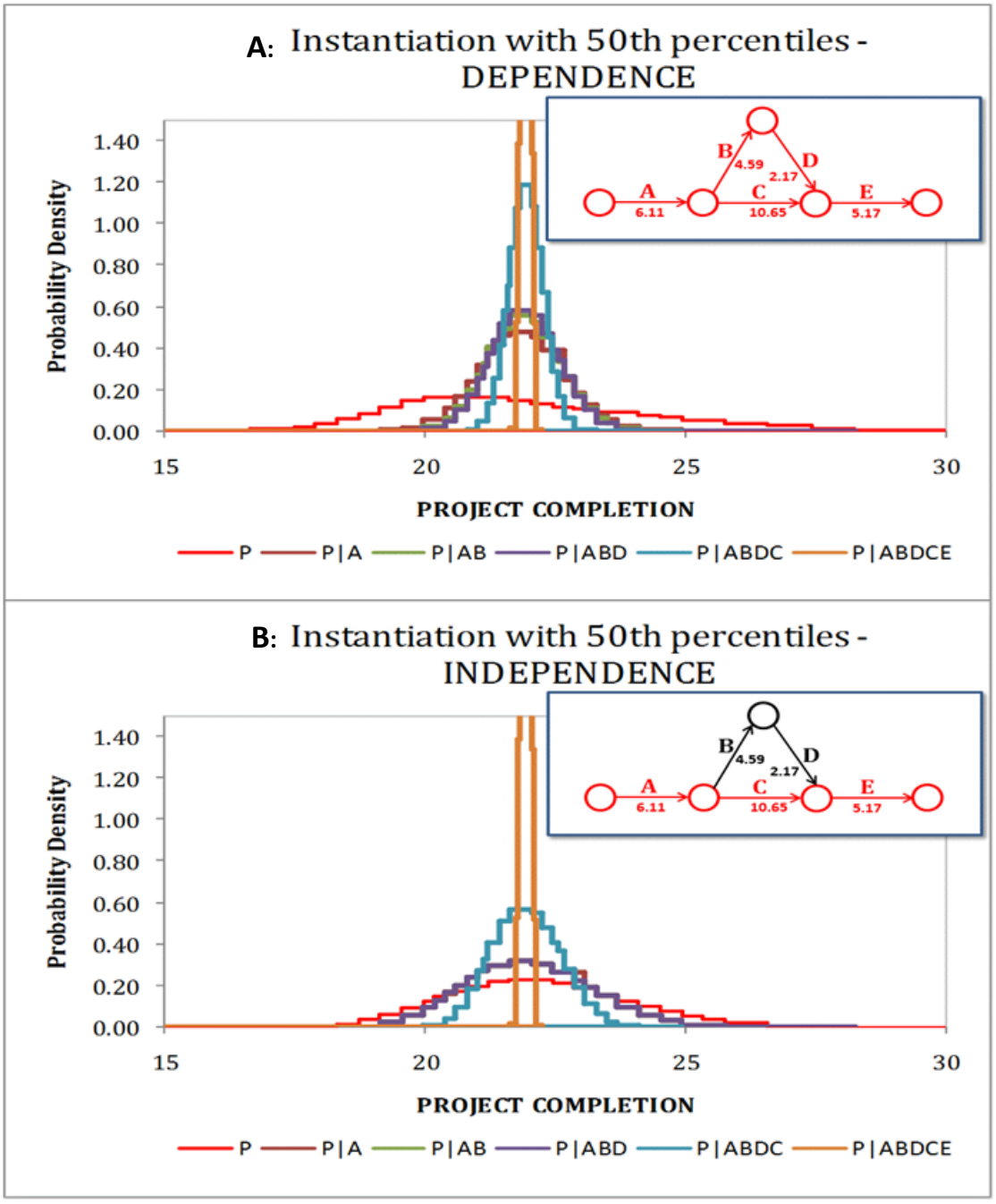


Figure 3-13: Distribution of Project Completion Times (given evidence of 50th percentile values of activity duration estimate across the project network)

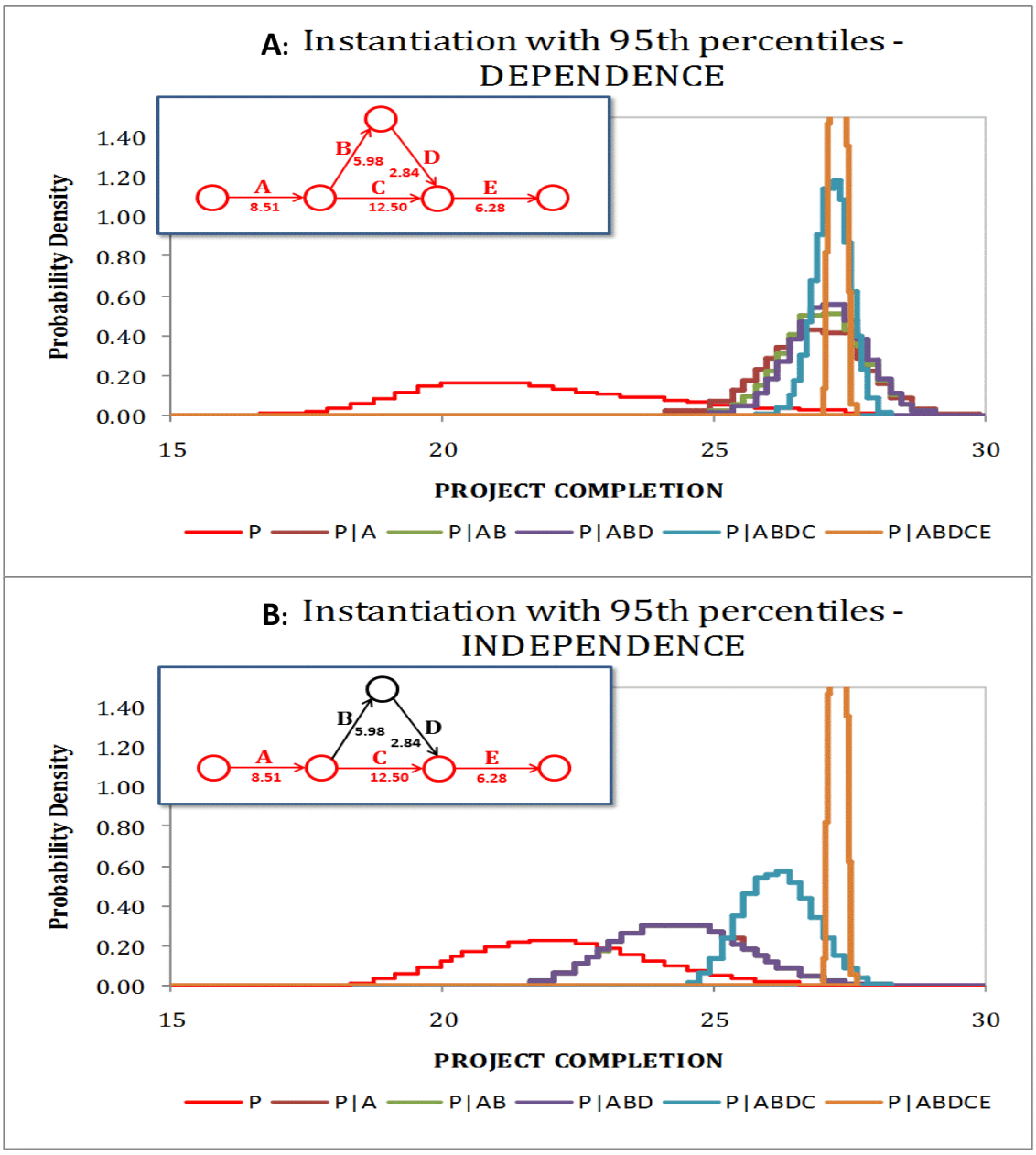


Figure 3-14: Distribution of Project Completion Times (given evidence of 50th percentile values of activity duration estimate across the project network)

3.9 Summary of *Method 1*

We have demonstrated in *Method 1* the importance of incorporating statistical dependence as it provides insight in probabilistic schedule analysis compared to the

situation where independence among the marginal distributions of activity durations is assumed. Very importantly, *Method 1* which has been described herein provides the unique opportunity of defining the shape parameter, *n*, for each project activity. This is in contrast with the method discussed in Chapter 4 where the modified PERT variance is applied and a single shape parameter, *n*, is defined and shared among all project activities. With reference to the procedural steps for *Method 1* which is discussed in Section 3.2 through Section 3.5, we provide below in Figure 3-15 a graphical summary that provides a snapshot insight into the general procedure described in developing the Bayesian Network Activity Dependence Model for Project Risk Analysis for *Method 1*.

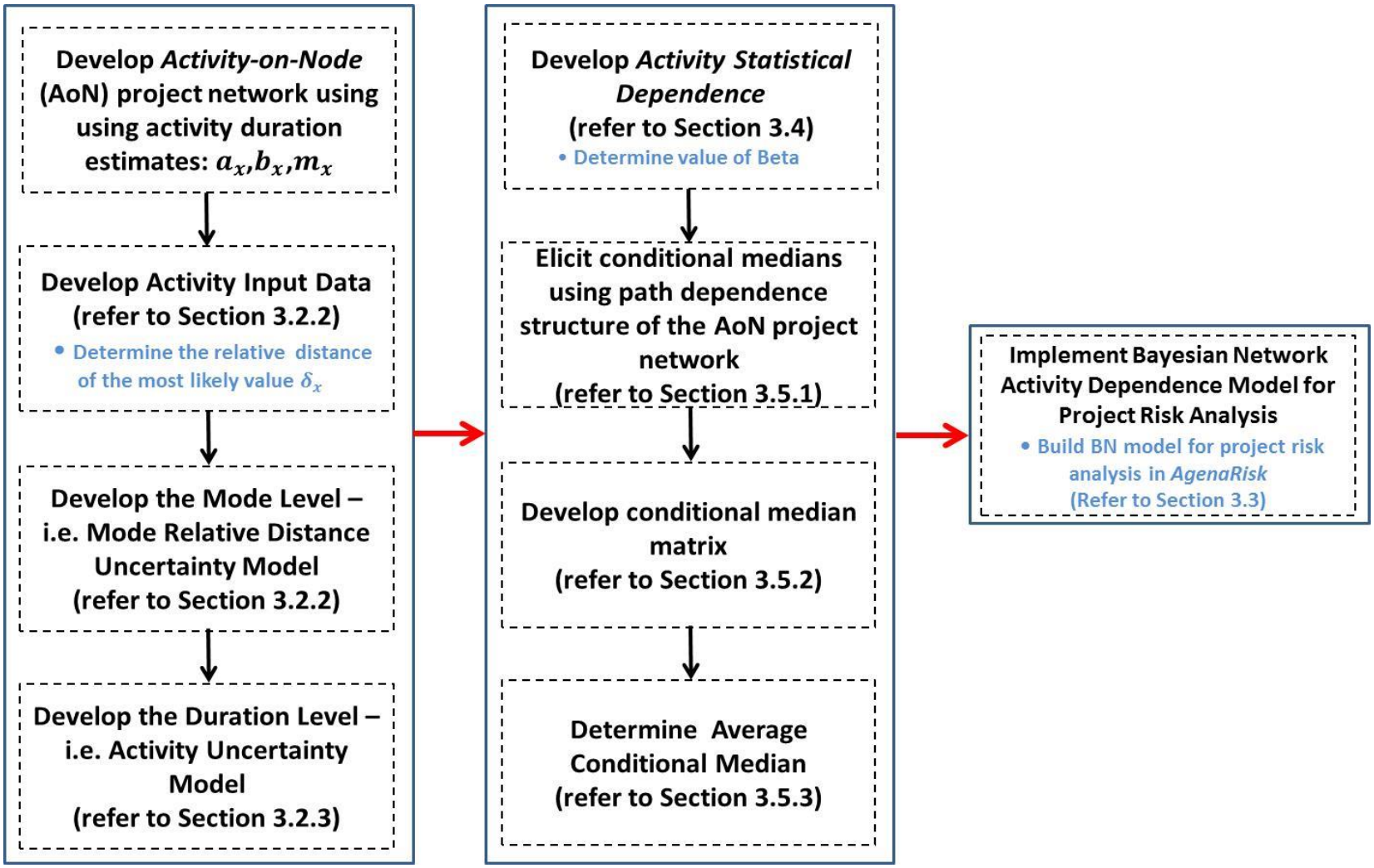


Figure 3-15 Graphical summary of the procedure used in developing BN Model for Method 1

4 Bayesian Network Activity Dependence Model for Project Risk Analysis using Modified PERT Variance

We propose the use of the modified PERT variance (Herrerias-Velasco et al., 2011), which in part is used in the quantitative representation and construction of the Bayesian Network model of *Method 2*. Our motivation stems from the viewpoint that the modified PERT variance addresses the limitation of classical PERT discussed by Hahn (2008) regarding the constant PERT variance assumption conditional on the range of support of the duration of a project activity. Thus:

$$\text{Modified PERT VARIANCE, } \mathrm{V}[\mathrm{X}|a,m,b] = C(\delta) \times \frac{(b-a)^2}{36} \tag{4-1}$$

where δ is given by Equation (3-1) and $C(\delta)$ is interpreted as the *PERT variance adjustment factor*, thus:

$$C(\delta) = \frac{5}{7} + \frac{16}{7} \times \delta(1-\delta) \tag{4-2}$$

Application area for the Modified PERT Variance is discussed in detail using *Case study 2*.

4.1 Case Study 2

Case Study 2 is descriptive of a shipbuilding project with 18 activities shown in Figure 4-1 as an Activity-on-Arrow (*AoA*) project network. The case study represents a

shipbuilding project which is typical of a project environment that is susceptible to the impacts of a risk factor with potential effects on project completion time.

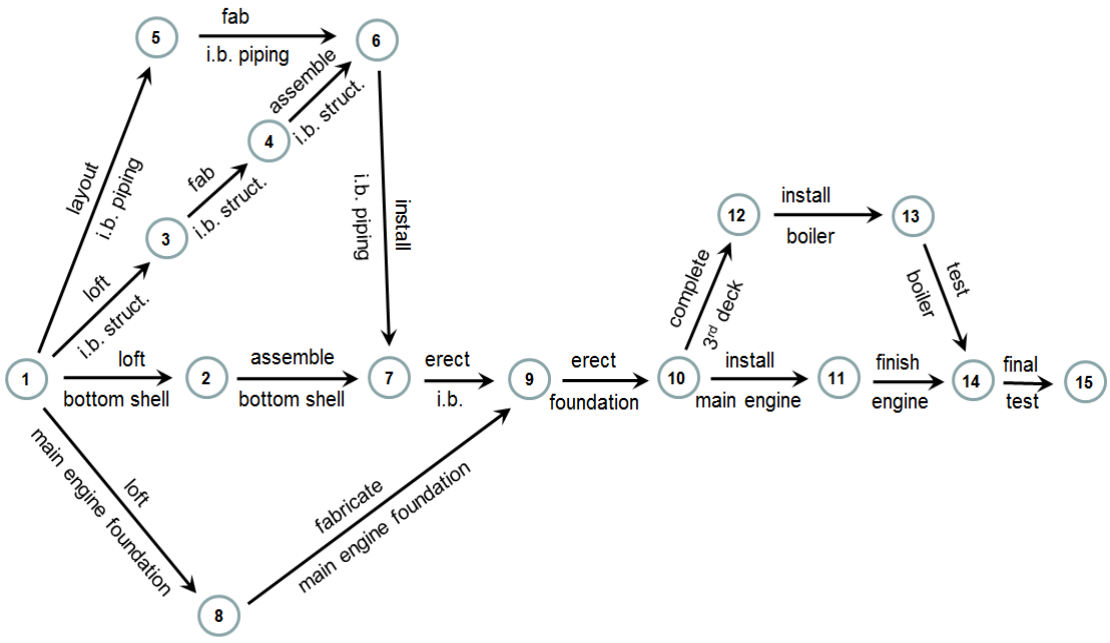


Figure 4-1: Example project for Case Study 2 (Taggart, 1980)

A representation of the example project network of Figure 4-1 is shown in Figure 4-2 as an Activity-on-Node (*AoN*) project network. It is important to point out the transformation of the *AoA* example project network of Figure 4-1 into the *AoN* example project network of Figure 4-2, thus: the Activity IDs displayed in the second column of Table 4-1 are used in developing the example project network of Figure 4-1 into an Activity-on-Node project network (see Figure 4-2).

The BN model for *Method 2* is represented (in Figure 4-3) and implemented in *AgenaRisk*® (Figure 4-4) for the *AoN* example project of Case Study 2 (see Figure 4-2).

Input Data: In Table 4-1 we provide relevant information for all the activities in Figure 4-1, including computations for the modified PERT variance and PERT Variance Adjustment Factor, which would be used in the analysis of Case Study 2.

Table 4-1: Case Study 2 Input Data

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

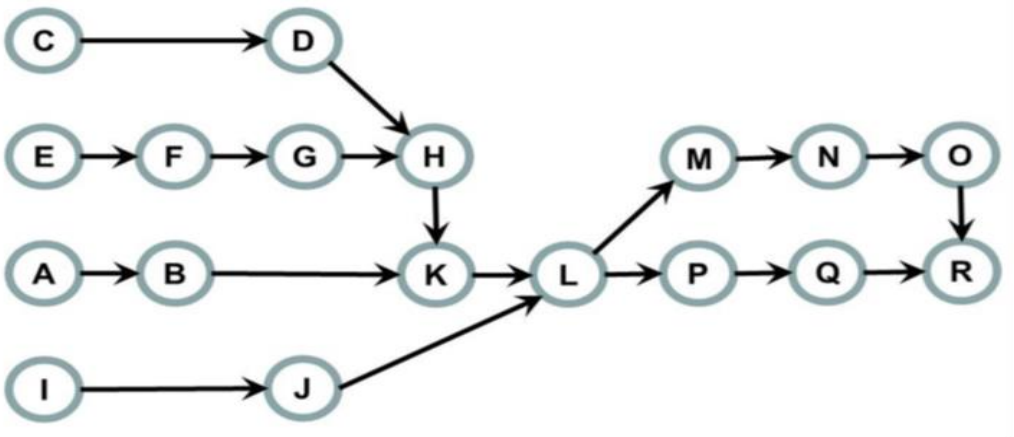


Figure *4-2*: *Example project for Case Study 2 (represented as Activity-on-Node project network)*

4.2 BN Model for Method 2

The BN model for *Method 2*, which is implemented for the 18-activity project network, is shown below:

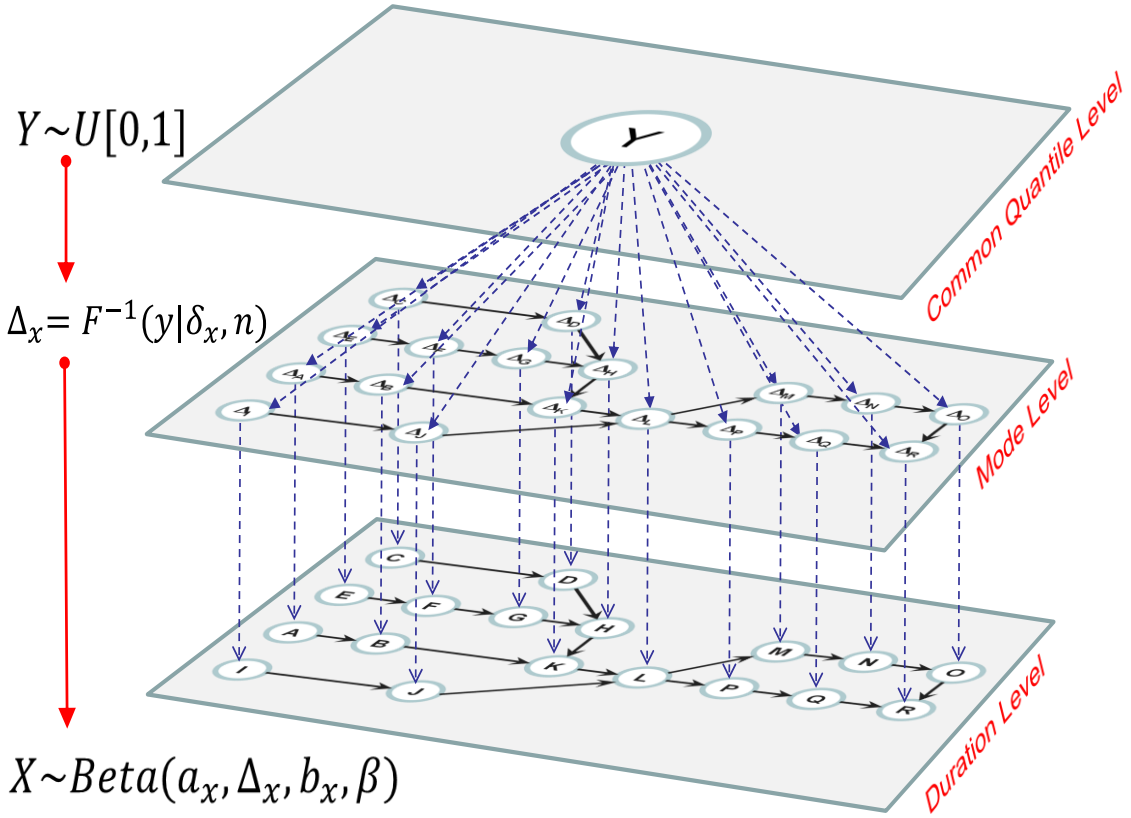


Figure 4-3: Bayesian Network Activity Dependence Model for Project Risk Analysis for Method 2 (illustrated for the example project of Case Study 2)

The elements or nodes of the BN model for *Method 2* are categorized into the Common Quantile Level, the Model Level, thus:

4.2.1 Common Quantile Level

Similar to the Common Quantile Level implemented in *Method 1*, we also utilize a non-informative distribution – the uniform distribution – to model the common quantile

level of *Method 2*, thus: $Y \sim U[0,1]$. The Common Quantile Level which is represented in Figure 4-4 by the *Y* node also serves as a monitoring variable that reflects the impact of a programmatic risk factor on the entire project.

4.2.2 Mode Level

The Mode Level in *Method 2* represents the Mode Relative Distance Uncertainty, modeled as a Two-Sided Power distribution (TSP) with parameters δ_x and n , thus: $\Delta_x \sim TSP(\delta_x, n)$:

$$f_{\Delta_x}(u|\delta_x, n) = n \times \begin{cases} \left(\frac{u}{\delta_x}\right)^{n-1} & 0 \le u \le \delta_x, \\ \left(\frac{1-u}{1-\delta_x}\right)^{n-1} & \delta_x \le u \le 1, \end{cases} \tag{4-3}$$

where Δ_x represents the uncertainty in the relative distance of the Most Likely Value to the lower bound *a* given the width of the support [*a*, *b*]. In the BN model for *Method 2* represented in Figure 4-4, the Model Level consists of 36 nodes. For each activity X we have nodes $\boldsymbol{D_X}$ and $\boldsymbol{LB_X}$ (where $X = A, B, C, \cdots, R$) to implement the quantile function of the TSP distribution in (4-3) given by:

$$F^{-1}(Y) = \begin{cases} a_x + \sqrt[n]{y(b_x - a_x)(m_x - a_x)^{n-1}}, & 0 \le y \le \delta_x, \\ b_x - \sqrt[n]{(1-y)(b_x - a_x)(b_x - m_x)^{n-1}}, & \delta_x \le y \le 1, \end{cases} \tag{4-4}$$

Recall from Equation (3-1) that $\delta_x = \frac{m_x - a_x}{b_x - a_x}$, which is the Relative Distance of the Most Likely Value of *X*. The node $\boldsymbol{LB_X}$ implements a Boolean node to determine, based on

the value of Y, if the lower or upper branch in (4-4) is selected. The node ***D_X*** evaluates $F^{-1}(Y)$ in (4-4)

4.2.3 Duration Level

The Duration Level represents activity uncertainty given the mode or the relative distance of the Most Likely Value, Δ_x, herein modeled as a Beta distribution, thus: $\Delta_x : X \sim \text{Beta}(a_x, b_x, \Delta_x, \beta)$. Applying the parameters $a_x, b_x, \Delta_x, \beta$ to the density function of the Beta distribution described in (van Dorp & Mazzuchi, 2004), we have:

$$g_X\left(x \middle| a_x, b_x, \Delta_x, \beta\right) \propto (x - a_x)^{\beta\alpha_x - 1}(b_x - x)^{\beta(1-\alpha_x)-1} \tag{4-5}$$

where α_x is:

$$\alpha_x = \Delta_x \times \frac{\beta-2}{\beta} + \frac{1}{\beta} \tag{4-6}$$

From the formulation presented from Equation (4-3) to (4-6), it should be noted that there are two parameters that are common to all project activities: *n* and β. The joint specification of these parameters will determine marginal average activity variance and average activity-to-activity statistical dependence.

In essence, the Duration Level implements the CPM evaluation of a project network's completion time given activity duration values, and which in Figure 4-4 consists of 54 nodes. Each activity X is represented by a Start node, a Finish node and Duration node as follows:

i. Start node: The *Start* node effectively represents the start of a project activity, and which is modeled as a continuous node using *Arithmetic Expression*, thus:

$$(\text{Start})_x = \text{Max}[(\text{Start})_y + (\text{Duration})_y; \text{ for all preceding activities y}] \quad (4\text{-}7)$$

ii. Finish node: The *Finish* node represents completion time or end of a project activity, and which is modeled as a continuous node using *Arithmetic Expression*, thus:

$$(\text{Finish})_x = (\text{Start})_x + (\text{Duration})_x \quad (4\text{-}8)$$

iii. Duration node: This represents the duration of a project activity, and which is modeled as a Beta distribution with parameterization defined by Equations (4-5) and (4-6).

4.3 AgenaRisk® representation of the BN model for *Method 2*

We show below in Figure 4-4 the implementation (in *AgenaRisk*®) of the BN Model for *Method 2* (representing the Dependence Scenario). Also, we show in Figure 4-5 the implementation of the example project for Case Study 2 (representing the Independence scenario) in *AgenaRisk*®.

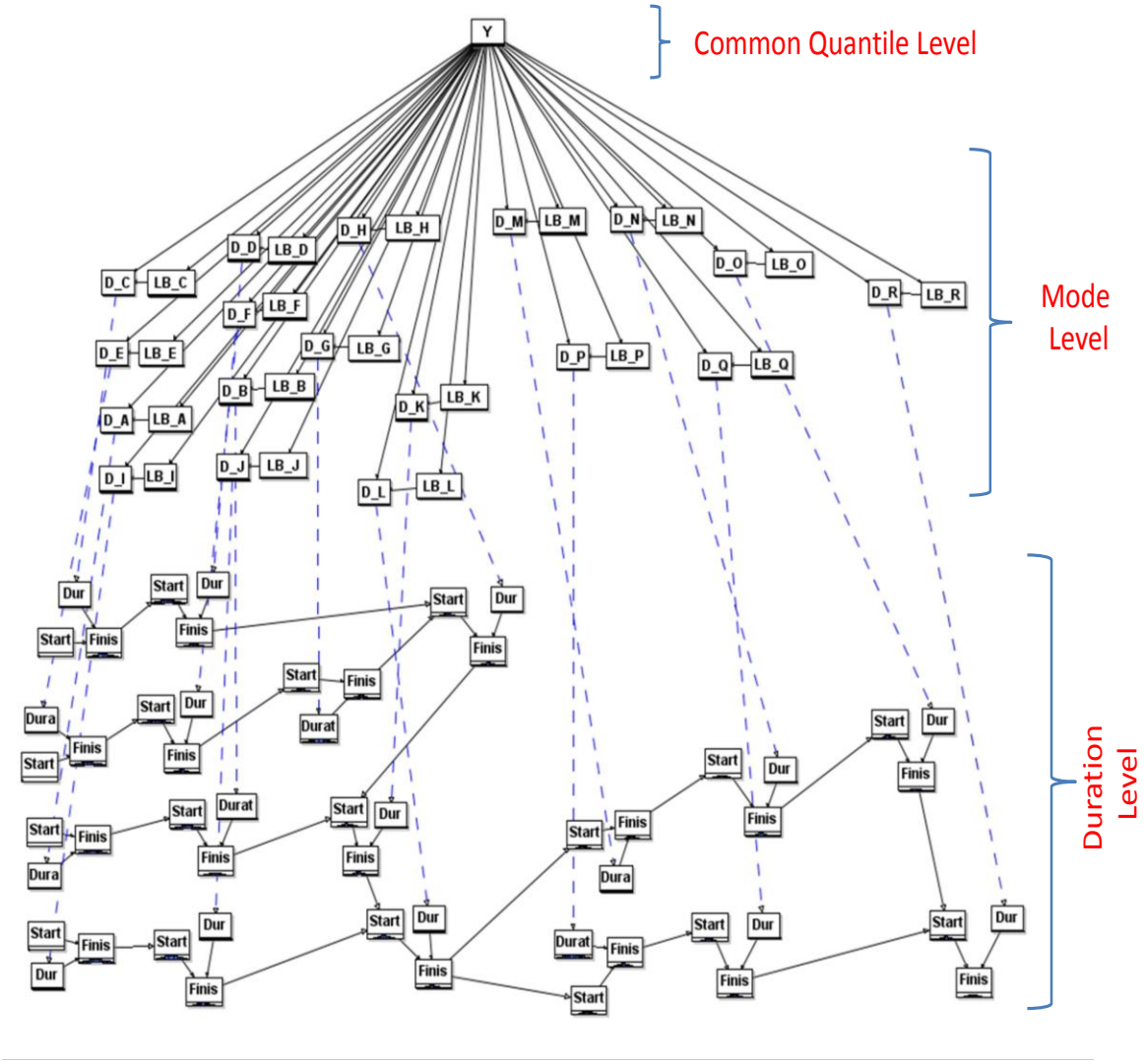


Figure 4-4: BN model for *Method 2* implemented in AgenaRisk® for the *example project of Case Study 2 (representing the Dependence scenario)*

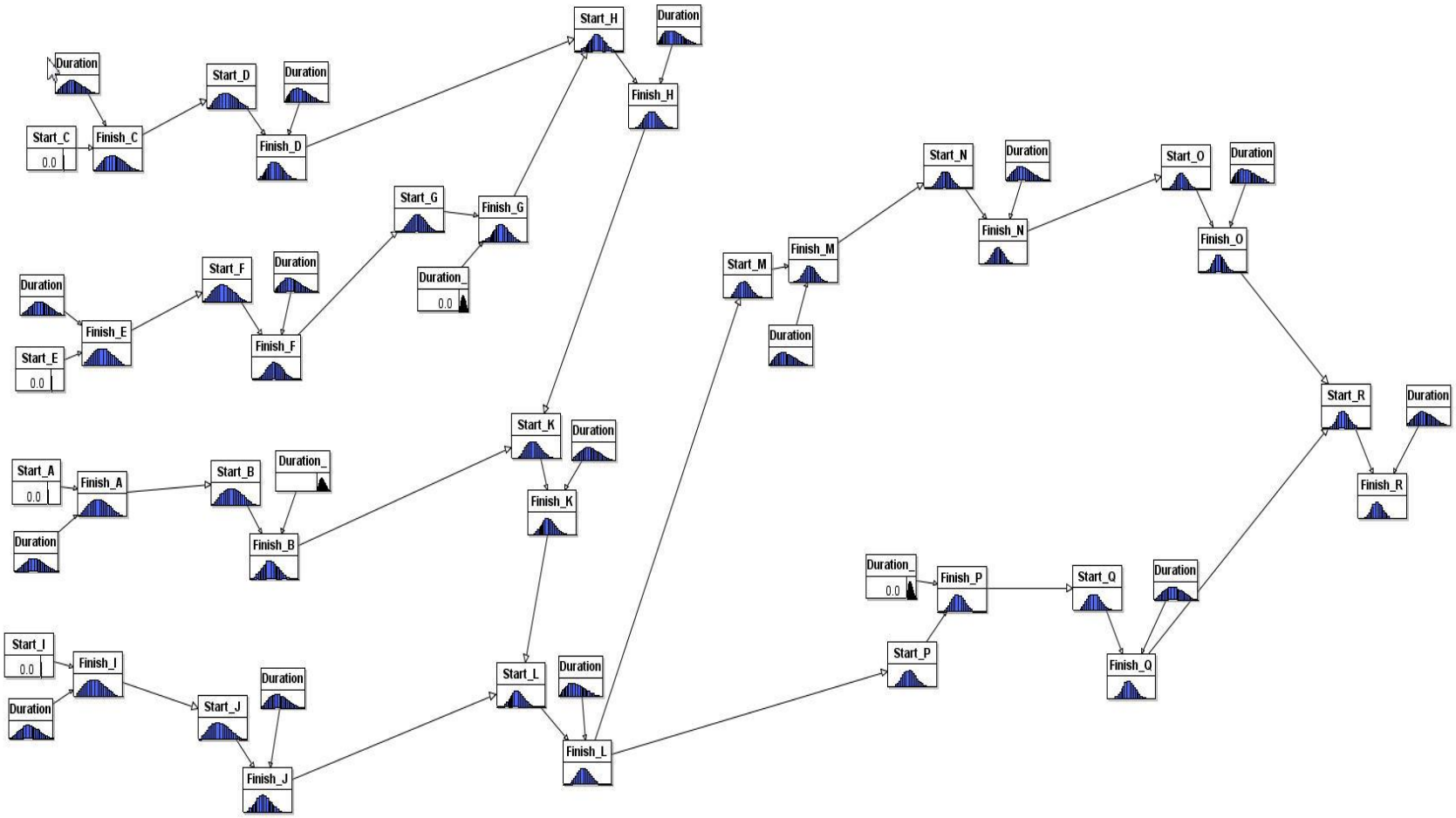


Figure 4-5: Example project network for Case Study 2 implemented in AgenaRisk (representing the Independence scenario)

4.4 Statistical Dependence Specification

Specification of 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 Bayesian Network with discrete nodes, Conditional Probability Tables (CPTs) are needed. Both, and other approaches toward constructing multivariate joint distributions, suffer from what is known to be "the curse of dimensionality". For instance, in the case of the example project network for Case Study 2 shown in Figure 4-1, a joint normal distribution approach towards statistical dependence would require the specification of $\binom{18}{2}$ = 153 additional correlations next to the activity duration data provided in Table 4-1. For a project network containing 100 activity durations this would grow to 4950 correlations that need to be specified.

Efficient modeling of continuous BN nodes in AgenaRisk® (Neil, Tailor, & Marques, 2007; Fenton & Neil, 2012) allows for dependence specification via parametric relationship between *parent* and *child* nodes reducing the need for dependence parameter specification considerably. In fact, the BN model for *Method 2* shown in Figure 4-3 only requires the specification of 2 additional parameters β and n in equations (4-3) and (4-5). However, this comes at a price in terms of flexibility of the dependence structures that can be accommodated. The BN model for *Method 2* shown in Figure 4-3 is also specifically developed, however, with an emphasis towards reducing the elicitation burden of statistical dependence specification among the activities in a PERT analysis to facilitate application of the approach.

4.5 Eliciting Activity-to-Activity Conditional Medians

Methods of elicitation of probabilities have been studied extensively (Garthwaite, Kadane, & O'Hagan, 2005). While suited for measuring the degree of dependence between two random variables, Pearson's product-moment correlation, Spearman's rank correlation or Kendall's tau do not lend themselves well for expert judgment elicitation. Despite having provided a decision analytic explanation of these measures (Kruskal, 1958), these 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 directly related to a conditional median of one of the random variables.

Therefore, for *Method 2* developed in this study, we suggest the elicitation of conditional medians between activity durations in the project network as an indirect elicitation procedure towards specification of parameters β and n in equations (4-3) and (4-5).

In essence, the elicitation protocol follows the set of responses from the question, thus: *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," knowledge of Activity A does not influence Activity B
- in (0.50, 1.0), knowledge of Activity A partially influences Activity B[2]
- " = 1.00," knowledge of Activity A determines Activity B

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

To further specify the degree of dependence in the BN model for *Method 2* (Figure 4-3), the elicitation of Activity-to-Activity conditional medians is suggested based on path-dependence structure of the example project network in Figure 4-2. Thus, we propose elicitation of Activity-to-Activity conditional medians along all of the eight paths of the project network in Figure 4-2. For example, for one of the paths (A, B, K, L, P, Q, and R) in the project network of Figure 4-2, we demonstrate the elicitation in such manner:

$$\Pr\{B > b_{0.5}|A > a_{0.5}\}, \Pr\{K > k_{0.5}|B > b_{0.5}\}, \Pr\{L > l_{0.5}|K > k_{0.5}\}, \qquad (4\text{-}9)$$
$$\Pr\{P > p_{0.5}|L > l_{0.5}\}, \Pr\{Q > q_{0.5}|P > p_{0.5}\}, \Pr\{R > r_{0.5}|Q > q_{0.5}\}$$

which completes the elicitation of Activity-to-Activity conditional medians along one of the eight paths (i.e. A, B, K, L, P, Q, and R) in Figure 4-2. 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! Recall the requirement of the specification of 153 correlations in the case that a joint multivariate normal distribution was proposed to capture statistical dependence among the 18 activities in the project network in Figure 4-1. To illustrate the mechanics of using conditional medians for *Method 2*, we use the values of the elicited 18 Activity-to-Activity conditional medians provided in the conditional median matrix in Figure 4-6. It is important to point out that the elicitation protocol described above is robust for the conditional median specifications shown in the conditional median matrix of Figure 4-6. The values of the conditional median matrix (shown, for example, in Figure 4-6) represent elicited conditional median based on the path-dependent structure of the

project network of Figure 4-2. By following the elicitation protocol discussed earlier in Section 4.5 and eliciting in a manner based on the path-dependence structure of the project network, the elicitation burden is reduced significantly.

	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 4-6: Conditional median matrix showing path-dependent activity-to-activity conditional medians for the example project in Figure 4-2

Observe from:

$$\Pr(Y > y_{0.5}|X > x_{0.5}) = \frac{\Pr(X>x_{0.5}|Y>y_{0.5})\Pr(Y>y_{0.5})}{\Pr(X>x_{0.5})} = \Pr(X > x_{0.5}|Y > y_{0.5}) \qquad (4\text{-}10)$$

that the conditional median matrix in Figure 4-6 is symmetric across the diagonal by definition. The upper (or lower) half above the diagonal lists the 18 Activity-to-Activity conditional definitions (or values).

We formulate the expression for the average conditional median of the example project network of Case Study 2, thus:

$$\overline{Med(\cdot|\cdot)} = \frac{\sum_{\substack{Y \in \{A,B,C,D,\ldots R\} \\ Y \neq X}} \sum_{X \in \{A,B,C,D,\ldots R\}} Med(Y \mid X)}{18 \times (18-1)} \tag{4-11}$$

Applying the values of the elicited conditional medians shown in Figure 4-6 to the expression in Equation 4-11, we have:

$$\overline{Med(\cdot|\cdot)} = \frac{\sum_{\substack{Y \in \{A,B,C,D,\ldots R\} \\ Y \neq X}} \sum_{X \in \{A,B,C,D,\ldots R\}} Med(Y \mid X)}{18 \times (18-1)} \approx 0.658$$

4.6 Uncertainty (*n*) and Dependence (β) parameter specification

The joint specification of *n* and Beta drives the marginal average activity variance and the average activity-to-activity statistical dependence. Given the project network structure implemented in the Bayesian Network model, we select the uncertainty (*n*) and dependence (β) parameters in such manner that the following conditions are met:

- Average conditional median elicited from experts is equal to the average conditional median defined by the Bayesian Network model.

- Average variance across the joint sample of all activities is equal to the average modified PERT variance across all of the activities.

Step-wise description of the procedure for selecting the parameters is shown below:

- Step 1: For different values of *n* and β, do the following:

    - Sample 1000 joint activity samples from joint prior (defined by BN)

- Evaluate Average Prior Marginal Activity Variance
- Evaluate Average Prior Activity Statistical Dependence

- Step 2: Select values of *n* and such that:
  - Average Prior Marginal Activity Variance close to Average Modified PERT Variance across project network
  - Average Prior Activity-to-Activity Statistical Dependence close to Elicited Average Activity-to-Activity Statistical Dependence

- Step 3: For selected values of *n* and implement the following:

  a. Sample 1000 Joint activities from Bayesian Network Prior

  b. Evaluate Average Prior Marginal Activity Variance

  c. Evaluate Average Prior Activity Statistical Dependence

  d. Repeat Steps a, b and c one 1000 times and plot:
  - Average Prior Marginal Activity Variance Uncertainty
  - Average Prior Activity-to-Activity Statistical Dependence Uncertainty

  e. Test if the Average Modified PERT Variance and elicited statistical dependence fall within 90% credibility limits of uncertainty distributions.

To generate the 1000 joint activity sample used in the analysis, the following steps are taken for each project activity: (1) sample a number from the uniform distribution and relative mode uncertainty, (2) With the relative mode uncertainty, solve for the parameters of the Beta distribution, and then sample from Beta distribution – the output thereof which represents a sample, and which is then performed 1000 times for each activity. From this joint sample, statistical measures of interest are evaluated, for example, the average prior activity-to-activity statistical dependence (i.e. prior average conditional median), etc.

In essence, the values of the uncertainty (*n*) and dependence (β) parameters are determined using the procedure described above in this section. The value of *n*, which is shared among the activities, is implemented in the Mode Level of the BN model represented in Figure 4-4. For example, the value of the parameter *n* is implemented as a constant within each of the *D_X* nodes (where $X = A, B, C, \cdots, R$) shown in Figure 4-4 (the AgenaRisk® representation of the BN model). Also, the determined value of dependence (β) parameter is implemented in the Duration Level (across each Duration node) of the BN model of Figure 4-4.

From the modified PERT variance column in Table 4-1 one evaluates an average variance of 1.79 and from the conditional medians in Figure 4-6 an average conditional median of approximately 0.66. Below we shall describe a numerical procedure to solve for the parameter n in (4-3) and the parameter β in (4-5) such that prior average variance

approximately equates to the value 1.69 and prior average conditional median approximately equates to the value 0.66 (van Dorp J. R., 2014).

Figure 4-7 was generated by sampling 1000 times (van Dorp J. R., 2014) from the prior distribution of the BN model in Figure 4-3 for the different values of β and n. Figure 4-7a displays the behavior of average activity variance as a function of β and n, whereas Figure 4-7b displays the behavior of average activity-to-activity conditional median. Observe from Figure 4-7a that average activity variances decrease with increasing values of β and n. On the other hand, one observes from Figure 4-7b an increasing (decreasing) behavior in the average activity-to-activity conditional median as the parameter β (parameter n) increases.

Figure 4-8 displays the data for Figure 4-7. The top part of Figure 4-8 displays in the shaded region values of the average activity variance within the range $(1.50,2.00)$. The bottom part of Figure 4-8 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 iso-contours, we select the values $n = 3.5$ and $\beta = 13$ by identifying cells with an average activity variance of 1.69 and an average activity-to-activity conditional median of 0.66 (van Dorp J. R., 2014).

Next, evaluating 1000 times the values of average activity variance and average activity-to-activity conditional median by sampling 1000 times from the prior distribution of the BN model in Figure 4-3, we evaluate remaining epistemic 90% probability interval (1.69,

1.83) of activity variance and remaining epistemic 90% probability interval (0.652, 0.671) for average activity-to-activity conditional median.

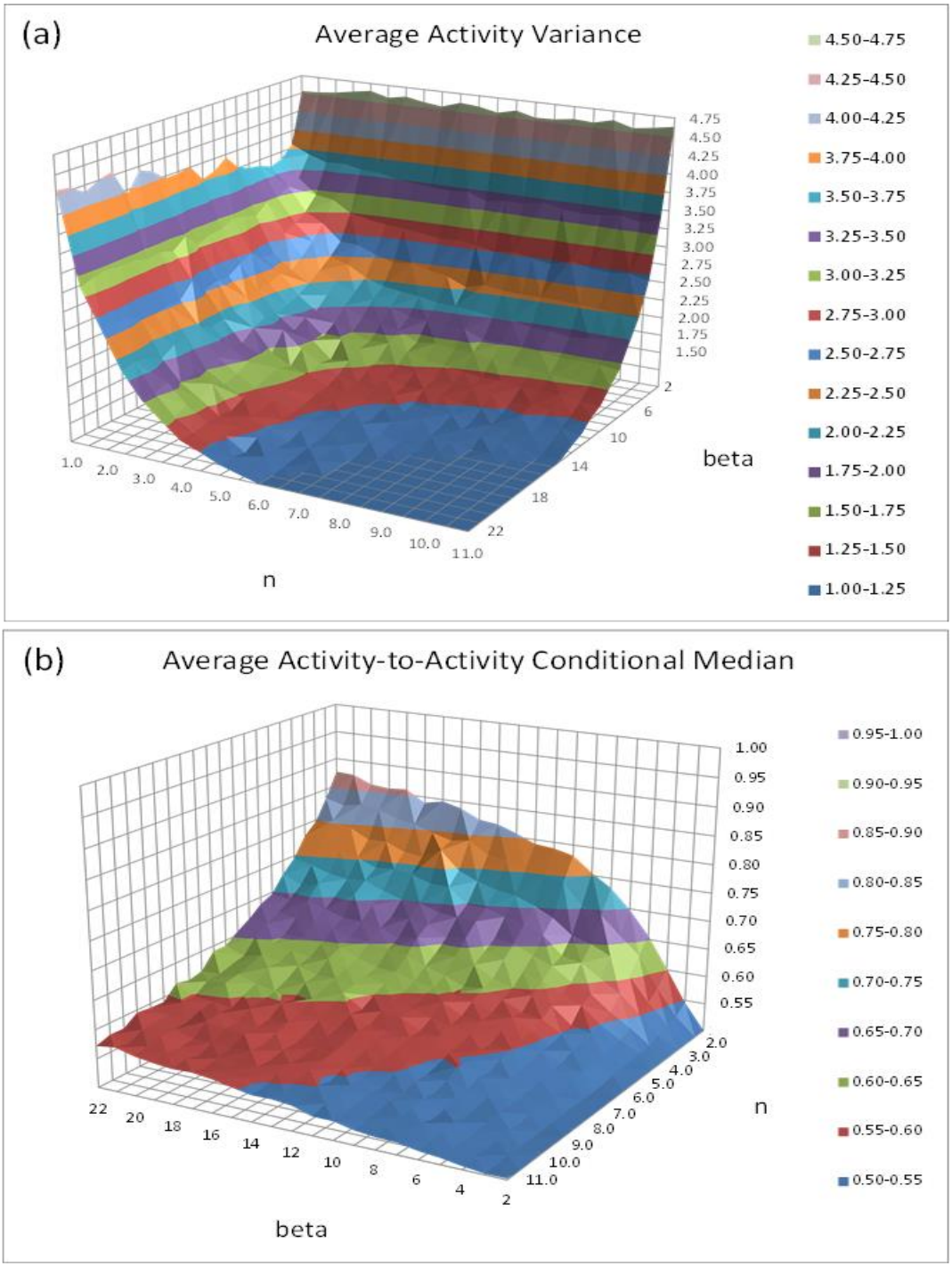
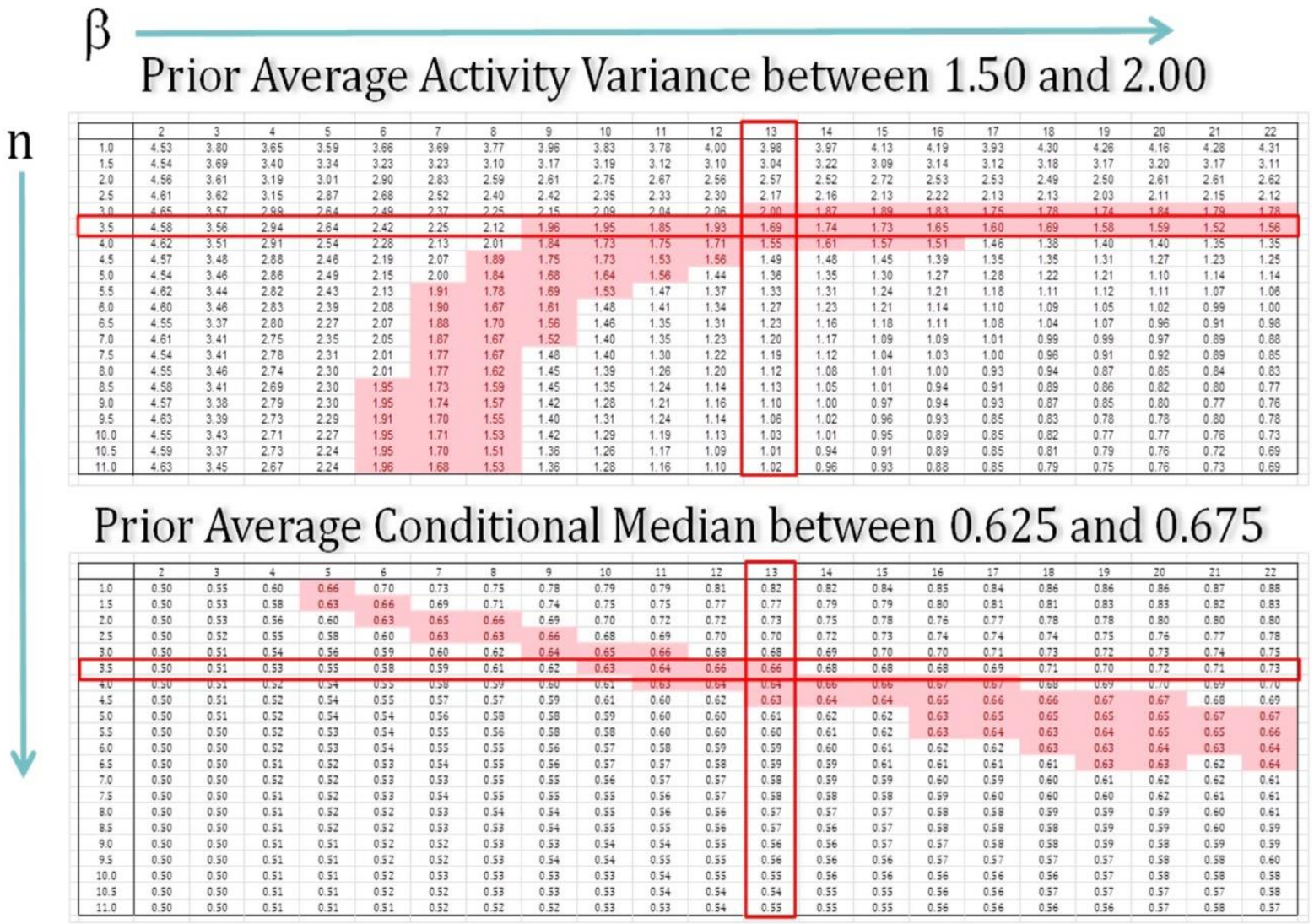


Figure 4-7: (a) Average Prior Marginal Activity Variance as a function $\boldsymbol{\beta}$ and $\boldsymbol{n}$ and (b) Average Prior Activity-to-Activity Conditional Median as a function $\boldsymbol{\beta}$ and $\boldsymbol{n}$



	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1.0	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
1.5	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
2.0	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
2.5	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
3.0	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
3.5	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
4.0	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
4.5	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
5.0	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
5.5	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
6.0	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
6.5	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
7.0	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
7.5	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
8.0	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
8.5	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
9.0	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
9.5	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
10.0	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
10.5	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
11.0	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

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1.0	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
1.5	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	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
2.5	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	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
3.5	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	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
4.5	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	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
5.5	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	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
6.5	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	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
7.5	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	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
8.5	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	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
9.5	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	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
10.5	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	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 4-8: Data for the 3D graphs in Figure 4-7

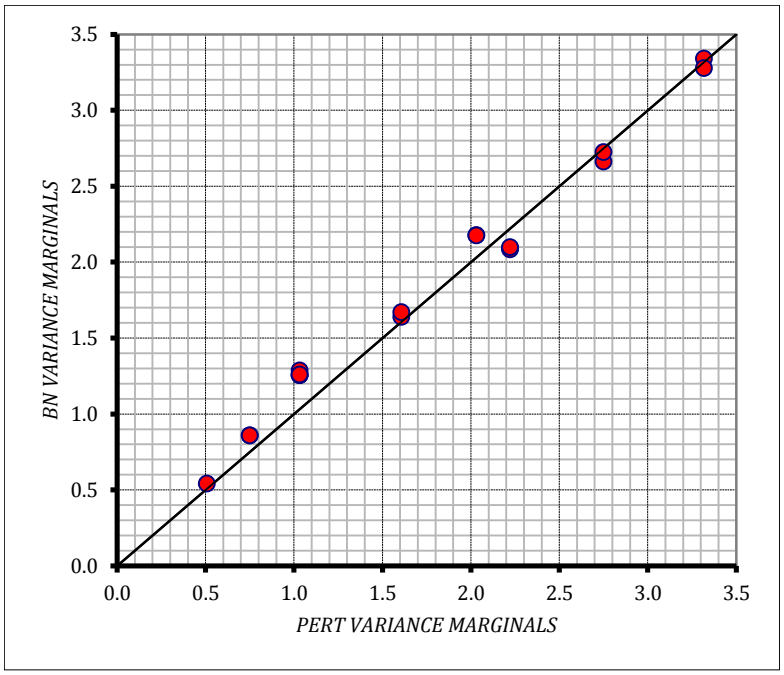


Figure 4-9: Visual comparison of BN activity variance and PERT modified activity variance for the eighteen activities listed in Table 1 (for values n=3.5 and β=13)

The first and second interval [(1.69,1.83) and (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 4-1, and 0.66 was evaluated from the conditional median matrix in Figure 4-6. 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 positive dependence.

Finally for verification, using the same prior sampling procedure, we evaluate activity variances of the BN model in Figure 4-3 using the values $n = 3.5$ and $\beta = 13$ and the data in Table 4-1. Figure 4-9 compares visually these variances with the modified activity

PERT variance provided in Table 4-1 demonstrating a close match across all eighteen activities.

4.7 Analysis of Case Study 2

We apply *Method 2* to the example project of Figure 4-1 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. To that end, we shall compare remaining project completion uncertainty at various milestones (or activity completions) using the BN model in Figure 4-3 (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 in which the assumption of statistical independence is implied among the marginal distributions of activity durations, is also implemented based on the example project of Figure 4-1. The analysis and implementation of this case of statistical independence in AgenaRisk®, however, include only the Duration Level elements of the BN model in Figure 4-3.

To allow for the aforementioned comparison, the prior marginal activity duration distributions were generated from the BN model in Figure 4-3 using the same sampling procedure as in Sections 4-4 and 4-6. Next, Beta distributions are fitted to the sampled prior activity distributions using Least Squares method implemented in an Excel spreadsheet for the project activity duration marginals (van Dorp & Nduka, 2013). Figure 4-10 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 4-10 that the Beta fitted distributions match well with prior sampled distributions of Activity A and Activity B. Also, P-P plots for the other 16 activity durations that appear as nodes in Figure 4-3 show similar behavior.

Table 4-2 lists the parameters of the Beta fitted marginal distributions of the activity durations in the BN model in Figure 4-3. In addition, the last row in Table 4-2 provides the 95th percentiles of these marginal distributions. In this example, the BN model is instantiated with these 95th percentiles representing activity completions. Subsequently, the state-of-the-art Bayesian inference algorithm implemented in AgenaRisk® enables propagation of influence across the nodes of the BN model, and also evaluates the remaining completion time distributions as the project progresses.

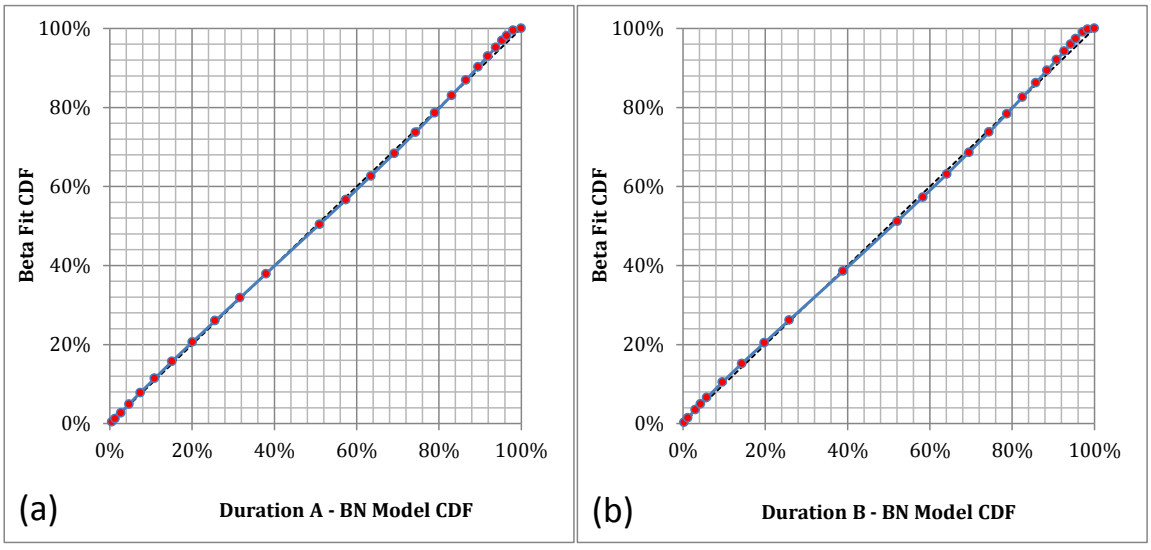


Figure 4-10: P-P plots comparing BN marginal distributions with Least Squares Beta fitted distributions for (a) Duration Activity A (b) Duration Activity B

Table 4-2: 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
95^{th}	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

Figure 4-11a and 4-11b depict the CPM evaluation of the example project in Figure 4-1 using the activity duration most likely values provided in Table 4-1 and the activity duration 95^{th} percentiles provided in Table 4-2. In addition, four intermediate milestones are identified in Figure 4-11. 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 4-2). 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 4-11.

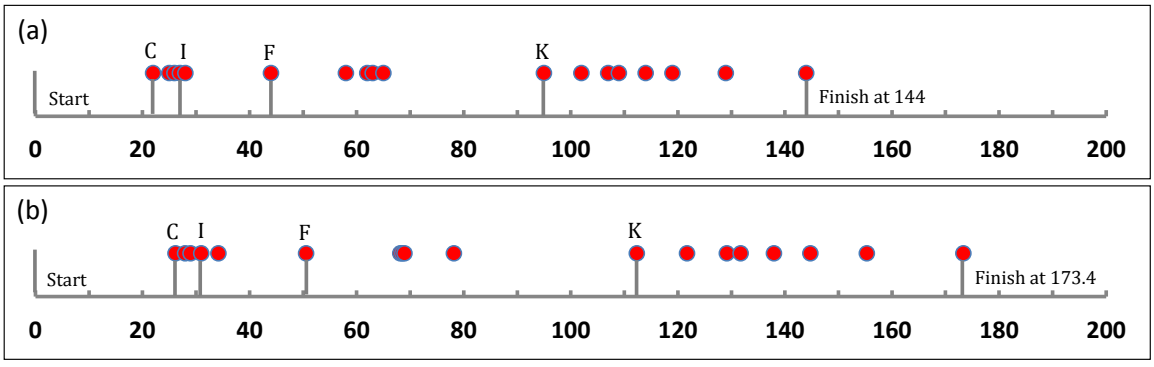


Figure 4-11: CPM evaluation of the example project in Figure 4-1 using - (a) activity duration most likely values shown in Table 4-1 (b) using 95th percentiles of activity durations shown in Table 4-2.

4.8 Prior Analysis of Project Completion Time Uncertainty

At the start of the project, we conduct analysis *a priori* using the example project network of Figure 4-1 and evaluating the scenario where statistical dependence is applied (Figure 4-3) and the scenario where statistical independence is implied (see the Figure 4-5 and introductory description of Section 4-7). We shall denote "DEP P" as the analysis scenario involving the BN model in Figure 4-3 (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 (Figure 4-5). In Bayesian analysis terminology these distributions are referred to as prior distributions.

Figure 4-12 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 4-12 is a result of the dynamic discretization methodology utilized in AgenaRisk® (Neil et al., (2007)). The project completion time at 173.4 (from Figure 4-11b) is indicated in Figure 4-12 using a vertical 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 comletion times evaluated under both scenarios exceed the CPM evaluated value of 144 days in Figure 4-11a. The "DEP P" pdf in Figure 4-12 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.

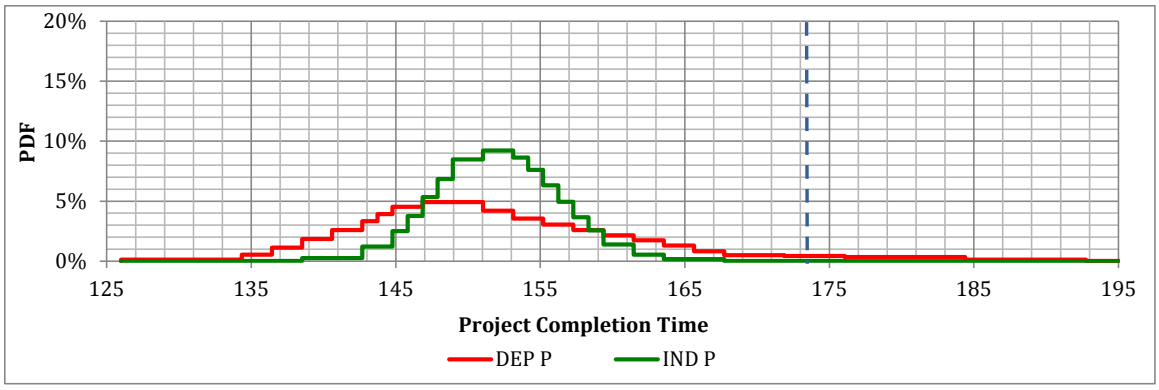


Figure 4-12: Prior project completion time distribution for the analysis of scenarios "DEP P" and "IND P"

4.9 Posterior Analysis: Monitoring 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 following the completion of Activity C. In Bayesian analysis terminology such a distribution is referred to as a posterior distribution.

It is important to also note that in conducting the posterior analysis using Method 2, the assumption implied previously in Section 3.8 is also applicable, that is, we instantiate at the Duration nodes of the BN model (both for the dependent and independent scenarios) under the assumption that "*if we had known at the start of the project the completion time of certain activities, what then would be the remaining completion time uncertainty as measured from the start of the project*."

Figure 4-13 displays the posterior distributions for the remaining completion time of the project at the four different milestones identified in Figure 4-11 (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 4-3.

Observe from Table 4-3 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, whereas no change is observed in the case of the independence scenario. Comparing Figure 4-12 with Figure 4-13a, it is observed that there is a shift to the right of the “DEP P” pdf and no shift observed for the “IND P” pdf.

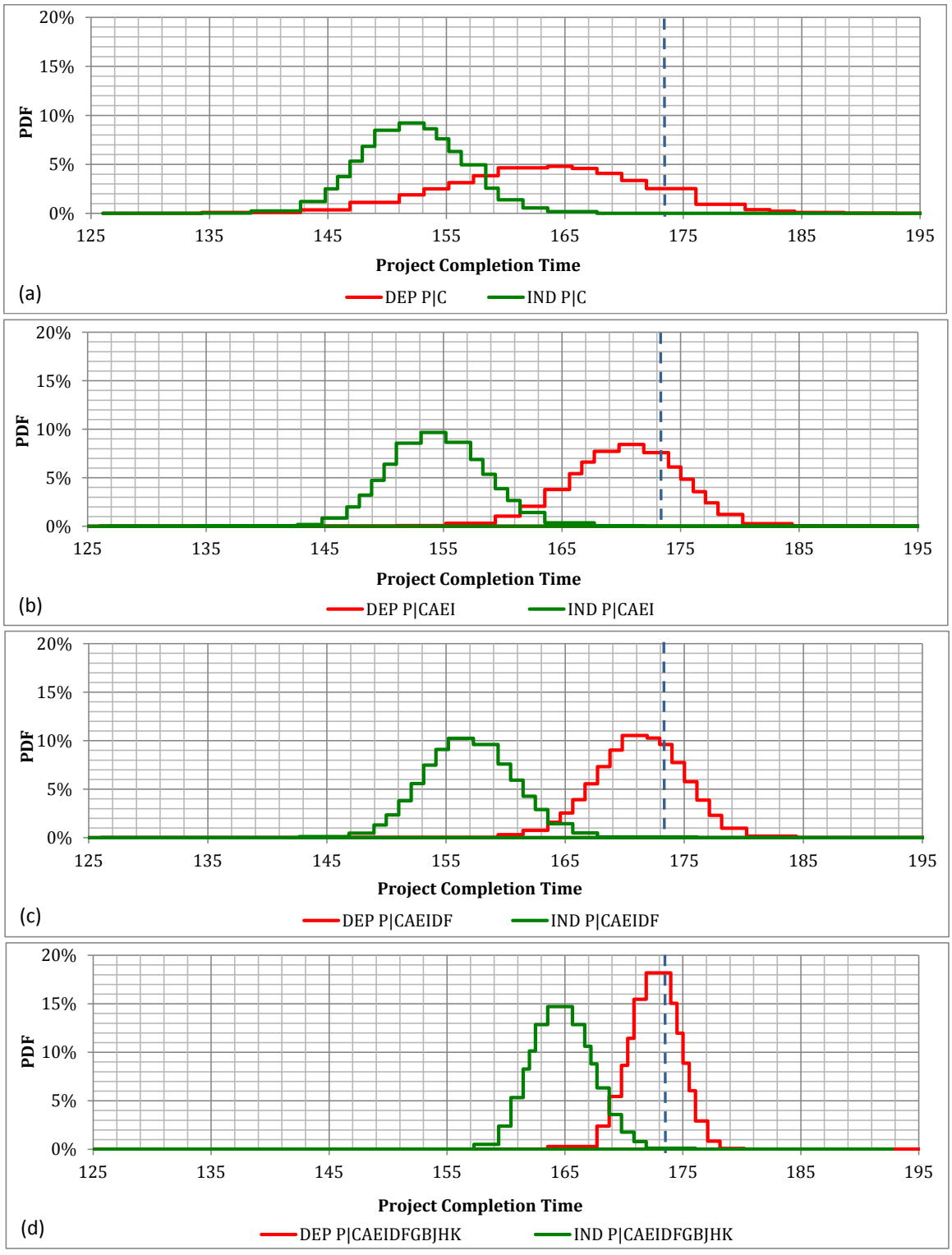


Figure 4-13: Posterior project completion time distributions for the analysis scenarios (a) "DEP P|C" and "IND P|C" (b) "DEP P|CAEI" and "IND P|CAEI" (c) "DEP P|CAEIDF" and "IND P|CAEIDF" (d) "DEP P|CAEIDFGBJHK" and "IND P|CAEIDFGBJHK"

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 4-12, and Figure 4-13a to Figure 4-13b 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 4-3.

Table 4-3: Mean values and standard deviations of remaining project completion time distributions evaluated at the project start and the four milestones identified in Figure 4-11 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

At the third milestone (the completion of Activity F), about one third into the project (see Figure 4-11), the degree of uncertainty in both the dependence and independence scenarios (reflected by their respective posterior distributions) have about evened out (Figure 4-13c). From the fourth row in Table 4-3, 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 in Figure 4-1.

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 4-11b a completion time of 173.4 days was evaluated for the entire project. 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 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 dependence among the activity uncertainty distributions. Observe from the last row in Table 4-3 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).

4.10 Summary of Method 2

Using *Method 2* we have demonstrated (via an illustrative example) the benefit of modeling statistical dependence among the activity durations in a project network on monitoring of remaining project completion time uncertainty. *Method 2* discussed

herein incorporates the modified PERT variance, and by extension, it leverages the advances previously made (Herrerias-Velasco et al., 2011) to address the constant PERT variance assumption conditional on activity's support. More so, one observes the advantages of *Method 2* in terms of increased precision of the analysis involved as well as in enhancements in the pace of learning about expected completion time of a project as it progresses. 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 avoiding costly schedule slippage.

In Table 4-4 shown below, we compare estimates of PERT means of activities with the mean values of activity durations of the BN model for the dependence scenario (Figure 4-4) and the BN model for the independence scenario (Figure 4-5). The means of activities shown in Table 4-4 for the dependence and independence scenarios are determined *a priori*. Insight learned from the Table 4-4 confirms the viewpoints earlier expressed about underestimation of completion times if statistical dependence is not incorporated in project schedule analysis. It clearly evident in Table 4-4 that the mean values of activity durations for the BN model for the dependence scenario are

consistently highest across all the activities of the project, while the PERT means of activities are the lowest across the project. More so, as described in the posterior analysis (Section 4.9), greater insight into activity and project completion times are learned more when statistical dependence is implied.

Table 4-5: Comparison of PERT means with BN means of project activities for Case Study 2

Activity Description	ID	a	m	b	PERT Mean	BN Mean for the Independence Scenario (a priori)	BN Mean for the Dependence Scenario (a priori)
Layout Bottom Shell	A	22	25	30	25.33	25.53	25.56
Assemble Bottom Shell	B	35	37	43	37.67	38.06	38.1
IB Piping Layout	C	19	22	29	22.67	23.06	23.104
IB Piping Fabricate	D	4	5	10	5.67	6.06	6.09
IB Struct. Layout	E	23	26	31	26.33	26.53	26.56
IB Struct. Fabricate	F	16	18	24	18.67	19.06	19.1
IB Struct. Assemble	G	11	14	20	14.50	14.8	14.83
IB Piping Install	H	6	7	12	7.67	8.09	8.09
Main Engine Found. Layout	I	25	28	33	28.33	28.53	28.56
Main Engine Found. Fabricate	J	33	35	40	35.50	35.8	35.83
Erect IB	K	27	30	37	30.67	31.06	31.1
Erect Foundation	L	6	7	11	7.50	7.79	7.82
Complete 3rd Deck	M	4	5	9	5.50	5.79	5.82
Install Boiler	N	6	7	10	7.33	7.53	7.55
Test Boiler	O	9	10	15	10.67	11.06	11.09
Install Main Engine	P	6	7	12	7.67	8.06	8.09
Finish Engine	Q	17	20	26	20.50	20.8	20.83
Final Test	R	13	15	20	15.50	15.8	15.83

Below in Figure 4-14, we provide a graphical summary that provides a snapshot insight into the general procedure described in this chapter for developing the Bayesian Network Activity Dependence Model for Project Risk Analysis for *Method 2.*

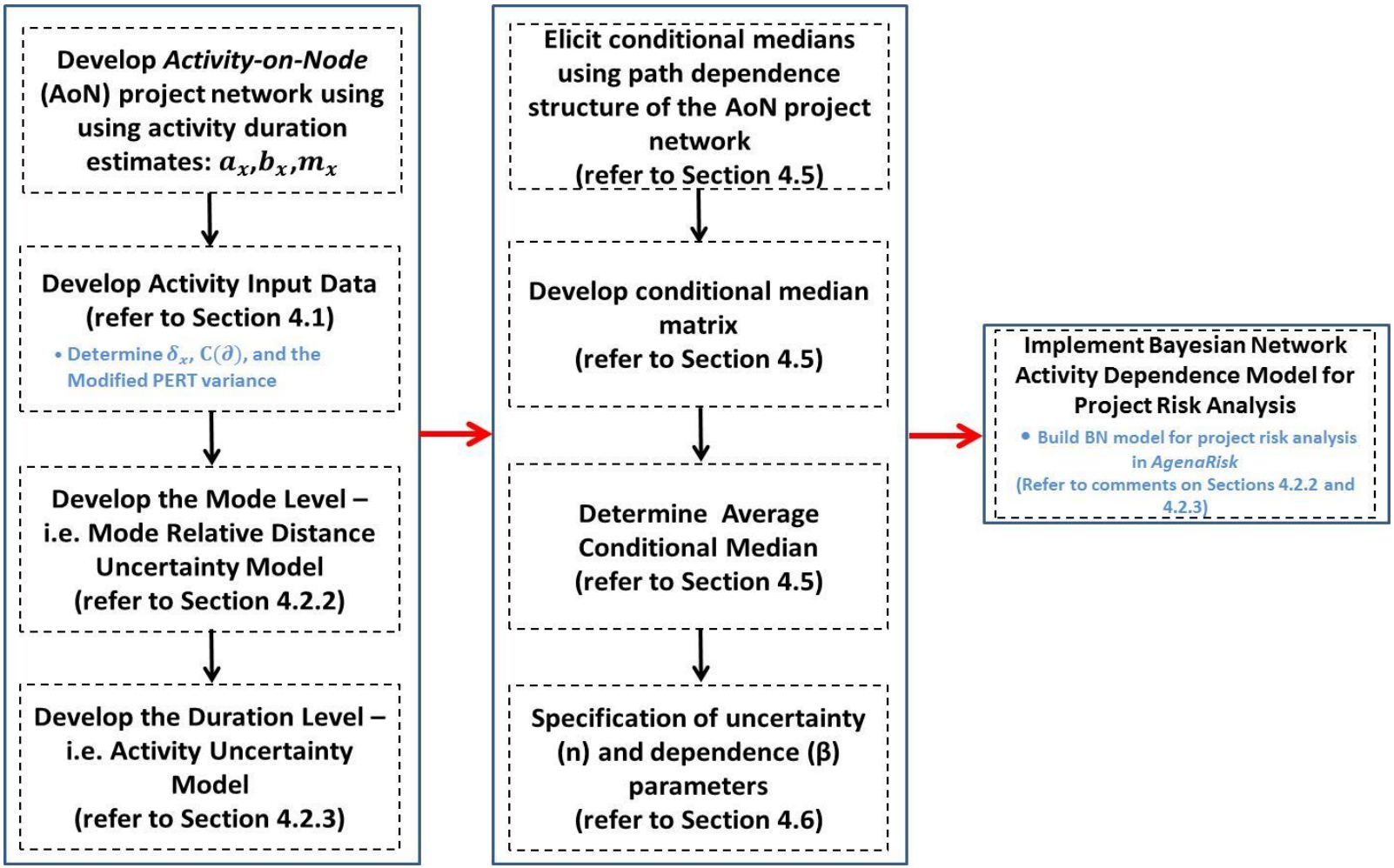


Figure 4-14: Graphical summary of the procedure used in developing BN Model for Method 2

5 Conclusions and Recommendation

5.1 Conclusions

Motivated by the capacity of Bayesian Networks to efficiently model epistemic uncertainty, we have developed two methods, described in this dissertation as *Method 1* and *Method 2*, which determine varying degrees of uncertainty in project completion times when a project is impacted by a programmatic risk factor. Both methods developed in this study address some of the inadequacies identified in some classical methods of project risk analysis, including: assumption of Statistical Independence, expert representation of uncertainty, and analysis of complex interrelationships between a risk factor and other project parameters.

In this study we have pushed the envelope by introducing a novel method based on Bayesian Networks for incorporating statistical dependence among activity durations and monitoring uncertainty in project completion times. Also, very importantly, we have introduced the use of Conditional Median as a measure for modeling dependence among activities in a project network. The speed of learning and insight derived about uncertainty in project completion times are greatly enhanced in utilizing both methods developed in this study. More so, by utilizing the methods discussed herein, the combined effect of propagation across the project network of completed (or partially completed) activities together with updating remaining activity uncertainty distributions provides more insight about the overall project schedule. From a project monitoring

standpoint, this would provide more accurate information about project performance, thereby enabling the project management team to apply corrective action(s) to the project, if need be.

5.2 Recommendation for Future Research

In this study, we investigated a common risk factor as a source of dependence. However, in reality, there could be instances where impacts to a project may result from more than one risk factor. Hence, modeling or incorporating more than one risk factor in the BN models developed in this study would be a natural extension of this study and a basis for future research.

Other areas of future research would involve: 1) investigating how extreme values of parameters *n* and β, and their interaction thereof, would affect the dependence structure and behavior of the BN model; 2) the use of Object-oriented Bayesian Network (OOBN) – a BN that provides means of modeling complex domains in terms of inter-related objects thereby making it reusable as a part of a larger BN model (Fenton & Neil, 2012; Koller & Avi, 1997). OOBN has potential promise in making the BN models developed herein more efficient especially for larger projects. Future research in this direction may entail developing efficient algorithm that would support forward and backward propagation of influence across the objects of OOBN; and 3) conducting a full pre-posterior analysis or predictive analysis comparison at various project milestones would be of interest, however, current computational tools are limited to conducting such analysis.

REFERENCES

Agena. (2012). Retrieved from http://www.agenarisk.com/

Al-Ahmad, W., Al-Fagih, K., Khanfar, K., Alsamara, K., Abuleil, S., & Abu-Salem, H. (2009). A Taxonomy of an IT Project Failure: Root Causes. *International Management Review, 5*(1), 93-106.

Alexander, C., & Sarabia, J. (2012). Quantile Uncertainty and Value-At-Risk Model Risk. *Risk Analysis, 32*(8).

Asllani, A., & Ettkin, L. (2007). An Entropy-Based Approach For Measuring Project Uncertainty. *Academy of Information and Management Sciences Journal*, 31-45.

Badiru, A. B. (1991). A simulation approach to PERT network analysis. *Simulation, 57*(4), 245-255.

Bayarri, M., & Berger, J. (2004). The Interplay of Bayesian and Frequentist Analysis. *Statistical Science, 19*(1), 58-80.

Bayes, T. (1763). Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, 53*, 370-418.

Bedford, T., & Cooke, R. (2001). *Probabilistic Risk Analysis: Foundations and Methods.* Cambridge: Cambridge University Press.

Bernardo, J. (2003). *Bayesian Statistics.* Oxford University Press.

Bidyuk, P., Terent'ev, A., & Gasanov, A. (2005). Construction and Methods of Learning of Bayesian Networks. *Cybernetics and Systems Analysis, 41*(4), 587-598.

Blomquist, N. (1950). On a measure of dependence between two random variables. *Annals of Mathematical Statistics, 21*(4), 593-600.

Bonnal, P., Gourc, D., & Lacoste, G. (2004). Where Do We Stand with Fuzzy Project Scheduling? *Journal of Construction Engineering and Management, 130*(1), 114-123.

Briggs, A. H. (1999). A Bayesian Approach to Stochastic Cost-Effectivness Analysis. *Journal of Health Economics, 8*(3), 257-261.

Callahan, M. T., Quackenbush, D. G., & Rowings, J. E. (1992). *Construction Project Scheduling.* New York: McGraw-Hill.

Chaloner, K., & Duncan, G. (1983). Assessment of a Beta Prior Distribution: PM Elicitation. *The Statistician, 32*(1/2), 174-181.

Chaloner, K., & Duncan, G. (1987). Some Properties of the Dirichlet-Multinomial Distribution and Its use in Prior Elicitation. *Communications in Statistics, Theory and Methods, 16*(2), 511-523.

Chanas, S., & Kamburowski, J. (1981). THE USE OF FUZZY VARIABLES IN PERT. *Fuzzy Sets and Systems, 5*(1), 11-19.

Chapman, C., & Ward, S. (2000). Estimation and evaluation of uncertainty: A minimalist first pass approach. *International Journal of Project Management, 18*(6), 369-383.

Cho, S. (2009). A linear Bayesian stochastic approximation to update project duration estimates. *European Journal of Operational Research, 196*(2), 585-593.

Cleden, D. (2009). *Managing Project Uncertainty.* Farnham: Gower Publishing Ltd.

Cogan, B. (2012). *Systems Engineering – Practice and Theory.* Rijeka: InTech.

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

Cottrell, W. D. (1999). SIMPLIFIED PROGRAM EVALUATION AND REVIEW AND REVIEW TECHNIQUE (PERT). *Journal of Construction Engineering and Management, 125*(1), 16-22.

Covaliu, Z., & Soyer, R. (1996). Bayesian Project Management. *Proceedings of the ASA Section on Bayesian Statistical Science* (pp. 208-213). American Statistical Association.

Covaliu, Z., & Soyer, R. (1997). Bayesian Learning in Project Management Networks. *Proceedings of the ASA Section on Bayesian Statistical Science* (pp. 257-260). American Statistical Association.

Cowell, R., Verrall, R., & Yoon, Y. (2007). Modeling Operational Risk with Bayesian Network. *The Journal of Risk and Insurance, 74*(4), 795-827.

Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Network.* New York: Cambridge University Press.

Demirer, R., Mau, R. R., & Shenoy, C. (2006). Bayesian Networks: A Decision Tool to Improve Portfolio Risk Analysis. *Journal of Applied Finance, 2*, 106-119.

Deshpande, G. (2012, May 15). Project as a Complex System - Feedback loops. Mindtree . Retrieved from http://www.blogs.mindtree.com/complex-system-feedback-loops

Dillon, R. L., Pate-Cornell, M. E., & Guikema, S. D. (2003). Programmatic risk analysis for critical engineering systems under tight resource constraints. *Operations Research, 51*(3), 354.

Douglas, D. E. (1978). PERT and Simulation. *Winter Simulation Conference. 1*, pp. 89-98. Piscataway: IEEE Press.

Druzdzel, M., & vanderGaag, L. (1995). Elicitation of Probabilities for Belief Networks: Combining qualitative and Quantitative Information. *Annual Conference on Uncertainty in Artificial Intelligence*, (pp. 141-148). Montreal.

Dubois, D., & Guyonnet, D. (2011). Risk-informed decision-making in the presence of epistemic uncertainty. *International Journal of General Systems, 140*(2), 145-167.

Duffey, M. R., & van Dorp, J. R. (1998). Project Risk Analysis Using Simulation of Activity Networks: Is It Valid, Practical, and Teachable. *American Society for Engineering Education annual Conference.* Seattle.

Fang, C., & Marle, F. (2011). A simulation-based risk network model for decision support in project risk management. *Decision Support Systems, 52*(3), 635-644.

Fenton, N., & Neil, M. (2012). *Risk Assessment and Decision Analysis with Bayesian Networks.* Boca Raton: CRC Press.

Finley, E., & Fisher, D. J. (1994). Project Scheduling Risk Assessment using Monte Carlo Methods. *Cost Engineering, 36*(10), 24.

Garthwaite, P., Kadane, J., & O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association, 100*(470), 680-700.

Goldratt, E. H. (1997). *Critical Chain.* Great Barrington: North River Press.

Grubbs, F. (1962). Attempts to validate certain PERT statistics or 'picking on PERT'. *Operations Research, 10*(6), 912-915.

Guiffrida, A. L., & Nagi, R. (1998). Fuzzy Set Theory Applications in Production Management Research: A Literature Survey. *Journal of Intelligent Manufacturing, 9*, 39-56.

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

Halpin, D. W., & Senior, B. A. (2010). *Construction Management.* Hoboken: John Wiley & Sons, Inc.

Helton, J., & Oberkampf, W. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety, 85*, 1-10.

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.

Hester, P. (2012). Epistemic Uncertainty Analysis: An Approach Using Expert Judgment and Evidential Credibility. *International Journal of Quality, Statistics, and Reliability, 2012*.

INCOSE. (2006). *Systems Engineering Handbook.* San Diego: International Council on Systems Engineering.

Intaver Institute Inc. (2005). *Event Chain Methodology in Project Management.* Calgary: Intaver Institute Inc.

Jensen, F. V. (2009). Bayesian Networks. *WIREs Computational Statistics, 1*(3), 307-315.

Khedr, M. (2006). Project Risk Management Using Monte Carlo Simulation. *AACE International Transactions*.

Khodakarami, V. (2009). *Applying Bayesian Networks to model uncertainty in project scheduling. (Doctoral dissertation).* Proquest.

Khodakarami, V., Fenton, N., & Neil, M. (2007). Project Scheduling: Improved Approach to Incorporate Uncertainty Using Bayesian Networks. *Project Management Journal, 38*(2), 39-49.

Kindinger, J. P. (2004). Programmatic Risk - Completing the Enterprise Picture. *National Academies of Science, Engineering, and Medicine's Enterprise Risk Management Workshop.* Washington, DC.

Kjaerulff, U. B., & Madsen, A. L. (2013). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis.* New York: Springer.

Koller, D., & Avi, P. (1997). Object-Oriented Bayesian Networks. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence* (pp. 302-313). Providence: Morgan Kaufmann.

Kotz, S., & van Dorp, J. R. (2004). *Beyond Beta - Other Continuous Distributions with Bounded Support and Applications.* Singapore: World Scientific Publishing Company.

Kotz, S., & van Dorp, J. R. (2010). Generalized Diagonal Band Copulae with Two-Sided Generating Densities. *Decision Analysis, 7*(2), 196-214.

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

Kulk, G., & Verhoef, C. (2008). Quantifying requirements volatility effects. *Science of Computer Programming, 72*(3), 136-175.

Leach, L. P. (1999). Critical Chain Project Management Improves Project Performance. *Project Management Journal, 30*(2), 39-51.

Lustig, H., Roussanov, N., & Verdelhan, A. (2011). Common risk factors in currency markets. *The Review of Financial Studies, 24*(11), 3731-3777.

MacCrimmon, K., & Ryavec, C. (1964). An analytical study of the PERT assumptions. *Operations Research, 12*(1), 16-37.

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

Meredith, J. R., & Mantel, S. J. (2006). *Project Manageent: A Managerial Approach.* Hoboken: John Wiley & Sons, Inc.

Meyer, A., Loch, C., & Pich, M. (2002). Managing Project Uncertainty: From Variation to Chaos. *MIT Sloan Management Review, 43*(2), 60-67.

Mo, J., & Zhao, Z. (2008). Using Hybrid Bayesian Networks to Model Dependent Project Scheduling Networks. *International Conference on Natural Computation*, *7*, pp. 163-166.

Moder, J., Phillips, C., & Davis, E. (1983). *Project Management with CPM, PERT and precedence diagramming.* New York: Van Nostrand Reinhold.

Na, K.-S., Li, X., Simpson, J. T., & Kim, K.-Y. (2004). Uncertainty profile and software project performance: A cross-national comparison. *The Journal of Systems and Software, 70*(1-2), 155-163.

Nasir, D., McCabe, B., & Hartono, L. (2003). Evaluating Risk in Construction Schedule Model (ERIC-S): Construction Schedule Risk Model. *JOURNAL OF CONSTRUCTION ENGINEERING AND MANAGEMENT, 129*(5), 518-527.

Neapolitan, R. E. (2012). *Probabilistic Reasoning In Expert Systems: Theory and Algorithms.* CreateSpace.

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

Ordóñez Arízaga, J. F. (2007). *A Methodology for Project Risk Analysis using Bayesian Belief Networks within a Monte Carlo Simulation Environment* . (Doctoral dissertation) Retrieved from ProQuest Dissertations and Theses Database. (UMI Number: 3260439).

Pate-Cornell, E. M. (1996). Uncertainties in risk analysis: Six levels of Treatment. *Reliability Engineering and System Safety, 54*(2-3), 95-111.

Pavlenko, T., & Chernyak, O. (2010). Credit Risk Modeling Using Bayesian Networks. *International Journal of Intelligent Systems, 25*(4), 326-344.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco: Morgan Kaufmann Publishers.

Perminova, O., Gustafsson, M., & Wikstrom, K. (2008). Defining Uncertainty in Projects. *International Journal of Project Management, 26*(1), 73-79.

Petrovan, A., & Costea, C. (2010). PERT and Fuzzy PERT Methods on Project Scheduling under Uncertainty. *Journal of Electronic and Computer Engineering, 3*(1), 99-102.

PMI. (2008). *A Guide To The Project Management Body Of Knowledge.* Newtown Square: Project Management Institute, Inc.

Ramgopal, M. (2003). Project Uncertainty Management. *Cost Engineering, 12*, 21.

Raz, T., Barnes, R., & Dvir, D. (2003). A Critical Look at Critical Chain Project Management. *Project Management Journal, 34*(4), 24-32.

Schutt, H. J. (1997). *Risk Management: Concepts and Guidance.* Darby: DIANE Publishing Co.

Szwed, P., van Dorp, J. R., Merrick, J., Mazzuchi, T., & Singh, A. (2006). A Bayesian Comparison Approach for Relative Accident Probability Assessment with Covariate Information. *European Journal of Operations Research, 169*(1), 157-177.

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

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

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

Trucco, P., Cagnoa, E., Ruggerib, F., & Grande, O. (2008). A Bayesian Belief Network Modelling of Organisational Factors in Risk: A Case Study in Maritime Transportation. *Reliability Engineering and System Safety, 93*(6), 823-834.

Van de Vonder, S., Demeulemeester, E., & Herroelen, W. (2007). A Classification of Predictive-Reactive Project Scheduling Procedures. *Journal of Scheduling, 10*(3), 195-207.

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

van Dorp, J. R. (2013, December 16). Dependent_Sampler_TSP_BETA_5.xlsm. Washington, DC, United States. Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

van Dorp, J. R. (2013, December 16). MedianMatrixEvaluator_5. Washington, DC, United States. Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

van Dorp, J. R. (2014, September 9). Behavior_n_beta_18.xlsx. Washington, DC, United States. Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

van Dorp, J. R. (2014, October 27). Dependent_Sampler_TSP_BETA_18.xls. Washington, DC, United States. Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

van Dorp, J. R., & Kotz, S. (2002a). The Standard Two-Sided Power Distribution and its Properties: With Applications in Financial Engineering. *The American Statistician, 56*(2), 90-99.

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

van Dorp, J. R., & Mazzuchi, T. (2004). Parameter Specification of the Beta Distribution and its Dirichlet Extensions Utilizing Quantiles. In A. K. Gupta, & S. Iadarajah, *Handbook of Beta Distribution and its Applications* (pp. 283-318). New York: Marcel Dekker, Inc.

van Dorp, J. R., & Nduka, I. C. (2013, December 12). MARGINALS_of_Activity_Nodes_DEPENDENCE_18.xls. Washington, DC, United States . Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

van Dorp, J. R., & Nduka, I. C. (2013, December 19). MARGINALS_of_Activity_Nodes_DEPENDENCE_5.xls. Washington, DC, United States. Retrieved from http://www.seas.gwu.edu/~dorpjr/PERT_BN.html

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.

Van Slyke, R. M. (1963). Monte Carlo Methods and the PERT Problem. *Operations Research, 11*(5), 839-860.

Viale, A. M., Kolari, J. W., & Fraser, D. R. (2009). Common Risk Factors in Bank Stocks. *Journal of Banking and Finance, 33*(3), 464-472.

Virine, L., & Trumper, M. (2007). *Project Decisions: The Art and Science.* Vienna: Management Concepts.

Wang, J., Li, J., Wang, Q., Zhang, H., & Wang, H. (2012). A Simulation Approach for Impact Analysis of Requirement Volatility Considering Dependency Change. *International Conference on Requirements Engineering: foundation for software quality* (pp. 59-76). Springer-Verlag, Berlin: Association for Computing Machinery (ACM).

Ward, S., & Chapman, C. (2003). Transforming Project Risk Management into Project Uncertainty Management. *International Journal of Project Management, 21*(2), 97-105.

Watt, D., & Willey, K. (2005). The complex, chaotic, and fractal nature of complex systems . *IEEE International Conference on Systems, Man and Cybernetics* (pp. 3155 - 3160). IEEE.

Weber, P., Medina-Oliva, G., Simon, C., & Iung, B. (2012). Overview on Bayesian networks Applications for Dependability, Risk Analysis and Maintenance areas. *Engineering Applications of Artificial Intelligence, 25*(4), 671-682.

Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Statistics.* Chichester: John Wiley & Sons.

Willams, T. (2004). Why Monte Carlo Simulations of Project Networks Can Mislead. *Project Management Journal, 35*(3), 53-61.

Yong, J., & Zhigang, Z. (2011). The project schedule management model based on the program evaluation and review technique and Bayesian networks. *IEEE International Conference on Automation and Logistics* , (pp. 379-383).

Zadeh, L. (1965). Fuzzy Sets. *Information and Control, 8*(3), 338-353.

Zhang, Q. (1989). A General Method Dealing with Correlations in Uncertainty Propagation in Fault Trees. *Reliability Engineering and System Safety, 26*(3), 231-247.

Zowghi, D., & Nurmuliani, N. (2002). A Study of the Impact of Requirements Volatility on Software. *Asia-Pacific Software Engineering Conference.* IEEE.

APPENDICES

Appendix A: Bayesian Network Formalism

We describe the theoretical underpinning of Bayesian Network as a paradigm for reasoning and decision making in problems that involve uncertainty. The information contained herein also describes some of the relevant syntax and semantics of BN, as well as the foundational knowledge for understanding the analysis and BN modeling approach described in Chapters 3 & 4. Also discussed in Appendix A are the unique functionalities of *AgenaRisk*® – especially the *dynamic discretization* of continuous nodes, and use of *expressions* and *partitioned expressions* – which enabled the modeling choices we used in building the framework for the Bayesian Network Activity Dependence Model for Project Risk Analysis.

A.1 Background of Bayesian Approach to Uncertainty Analysis

The Bayesian approach is premised on the description of probability as a conditional and rational measure of uncertainty (Bernardo, 2003). This approach to probability and statistics is helpful in understanding the concept of Bayesian Network and its associated learning techniques. In the following sub-section we provide a brief description of the Bayesian approach.

Probability theory provides a formal framework for reasoning about uncertain events, and of which there are two major paradigms used by mathematical statisticians: *frequentist* and *Bayesian.* The distinction between the two paradigms of statistical

inference is based on their respective definition and interpretation of probability. To the *frequentist*, probability is used in the context of interpreting the frequency by which events will occur over repeated trials. Probability in this perspective is considered as a description of the physical property of the system or as a long-run frequency of events. This is in contrast with the *Bayesian* view which uses probability to express belief in a statement about unknown parameters; hence with respect to an event, probability is interpreted as the person's *degree of belief* in that particular event.

The debate about the *frequentist* and *Bayesian* approach is renowned; however, further information on the two approaches can be found in (Briggs, 1999; Neapolitan, 2012; Bayarri & Berger, 2004). Both approaches have their advantages and disadvantages, but it is important to note that in situations where there is limited or no historical data to draw upon, the *frequentist* approach offers limited value in providing means by which uncertainty in a given system can be effectively quantified. In this instance, however, the Bayesian approach provides more robust means for quantifying uncertainty. For example, if we want to determine the impact on project duration as a result of a change in the technical requirement of a project to develop a novel and unique weapon system, it will be difficult to effectively model or ascertain the uncertainty in the project duration using data-driven *frequentist* approach to risk assessment since no such type of weapon system had been produced before (and therefore no relevant data to draw upon).

The Bayesian approach is rooted in the work of Thomas Bayes (Bayes, 1763). His work which was based on the mathematical theory of probability is the foundation of what is currently known as the Bayes' Theorem. The expression of the Bayes' Theorem provides the conditional probability of an event *A* given event *B*, and is given by:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad \text{(A-1)}$$

The Bayes' Theorem expressed above as *Equation A-1* describes a means of updating subjective degree of belief to account for evidence (where $P(A|B)$ = *posterior* probability; $P(B|A)$ = *likelihood function*; $P(A)$ is the *prior* probability; and $P(B)$ represents the marginal probability), and all of which form the basis of what is known as *Bayesian Inference*.

Typically, the Bayesian approach to a problem involves: the formulation of a model that captures the variables in a problem domain of interest; the development of *prior* distribution over the unknown variables of the model, which captures our beliefs about the problem domain before evidence is observed; and lastly, the application of Bayes' Theorem after some evidence has been observed to determine the *posterior* distribution.

A.2 Bayesian Network

Bayesian Network (also referred to as Bayesian Belief Network or Belief Nets) is a graphical model for encoding probabilistic relationships among variables in a problem

domain. They are efficient tools for performing inference and analysis, especially when limited information about the underlying problem domain is available (Bedford & Cooke, 2001). BNs compactly specify the joint probability distribution of the variables underlying a domain by means of a directed acyclic graph with nodes that represent random variables and arcs that represent causality and/or influences between adjacent nodes. Some of the fundamental properties of a BN include:

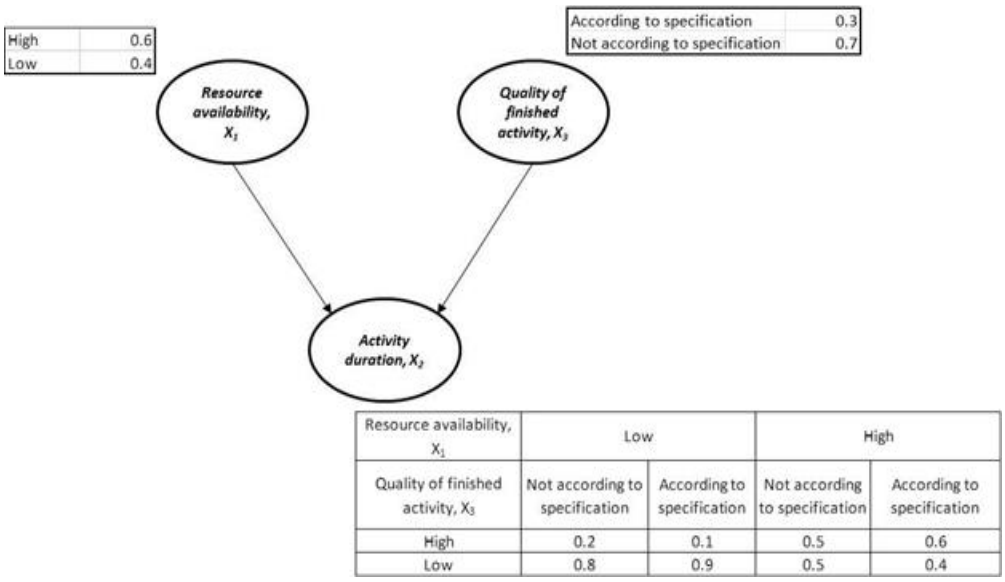
- Nodes in a BN can be discrete or continuous. For discrete nodes, there exists a finite set of *states* that consist of values of the variable that are mutually exclusive and exhaustive.
- Each node with parent node(s) is associated with a conditional probability table. For example, a node in a BN, say *A*, with parent node *pa(A),* has a conditional probability table $P[A \mid pa(A)]$ assigned to it (Jensen, 2009). Instances where a node, say *A*, does not have a parent, then absolute probabilities *P(A)* are used (Bidyuk, Terent'ev, & Gasanov, 2005). Conditional probabilities are expressed as Conditional Probability Tables (CPT) for discrete variables and Conditional Probability Distributions (CPD) for continuous variables.

Given a set of variables, $U = \{A_1, A_2, ..., A_n\}$, the compact representation of the joint probability distribution over all these variables in a Bayesian Network is achieved by means of the *Chain Rule,* which postulates that the full joint probability distribution over the variables of a Bayesian Network is given by the product of all the conditional probabilities specified in the BN:

$$P(U) = \prod_{i=1}^{n} P[A_i | pa(A_i)] \quad \text{(A-2)}$$

In essence, by using a Bayesian Network, the goal is to represent a joint probability distribution over some set of random variables that define a problem domain. Of importance to the theory that underpins graphical models (for example, Bayesian Networks) is the probability concept of conditional independence (Whittaker, 1990; Fenton & Neil, 2012; Darwiche, 2009). A set of conditional independence assumptions encoded by a BN structure ensures that the joint probability distribution over all variables is equal to the product of the conditional probabilities of each node given its parents.

To illustrate some BN concepts, for example, the arcs from node X_1 to X_2 and X_3 to X_2 shown in *Figure A-1* encode the underlying assumption that there exists direct probabilistic influence and/or causal relationship of X_1 on X_2 and X_3 on X_2. The nodes X_1 and X_3 are called the *parent* of X_2. Additionally, there is a Conditional Probability Table (CPT) that is associated with every discrete node of a Bayesian Network.



Resource availability, X_1	Low		High	
Quality of finished activity, X_3	Not according to specification	According to specification	Not according to specification	According to specification
High	0.2	0.1	0.5	0.6
Low	0.8	0.9	0.5	0.4

Figure A-1: Example BN model with three nodes

A.3 Node Independence Assumption and D – Separation

The notion of independence plays an important role in Bayesian Network modeling. The conditional independence assumptions encoded in a BN enables the compact factorization of the joint probability distribution of the variables underlying a domain. For example, consider the Bayesian Network of *Figure A-1*, a full joint probability distribution of the BN is represented as: $P(X_1, X_2, X_3) = P(X_1)P(X_2 \mid X_1, X_3)P(X_3)$

Given that the graphical structure of a BN represents a set of independencies that the probability distribution must satisfy, the expression of the full joint probability distribution shown above is an equivalent representation of the topology of the BN shown in *Figure A-1* and which also specifies that the assumptions of direct dependencies for the BN are fully encoded by the expression. In essence, the three

expressions to the right hand-side of the equation [i.e. $P(X_1)P(X_2|X_1,X_3)P(X_3)$] represent exactly the number of parameters or Node Probability Tables that will need to be defined for computation and probabilistic inference. If no assumption of independence is specified for the BN of *Figure A-1*, then a higher number of parameters must be specified for a robust probabilistic inference. In general, the number of parameters to be specified increases exponentially as the number of variables in a BN increases.

When a BN is constructed, it is important to ensure that the independent assumptions implied by the BN is consistent with beliefs about the underlying domain modeled by the BN. Propagation of evidence between nodes in a Bayesian Network is determined by the conditional dependency structure; therefore, to develop a Bayesian Network structure, it is imperative to take into consideration the different types of structural relationships that exist between variables of a BN.

Suppose a variable, *Y,* is directly linked to other variables (say *X* and *Z*) in a Bayesian Network. Generally, the point of interest is in determining under what structural condition of a BN will enable a node to be updated with evidence from another node. The concept of *d-separation* is an important formalism for encoding independencies in a Bayesian Network, as it plays a significant role in determining the path in a BN through which evidence is propagated. For instance, two nodes, say *X* and *Z*, are *d*-separated by a node *Y* if all paths between X and Z are blocked by a node *Y*. *Figure A-2* describes the

three general cases of connection that define conditional independence in a Bayesian Network, thus: *Serial*, *Diverging*, and *Converging* connections.

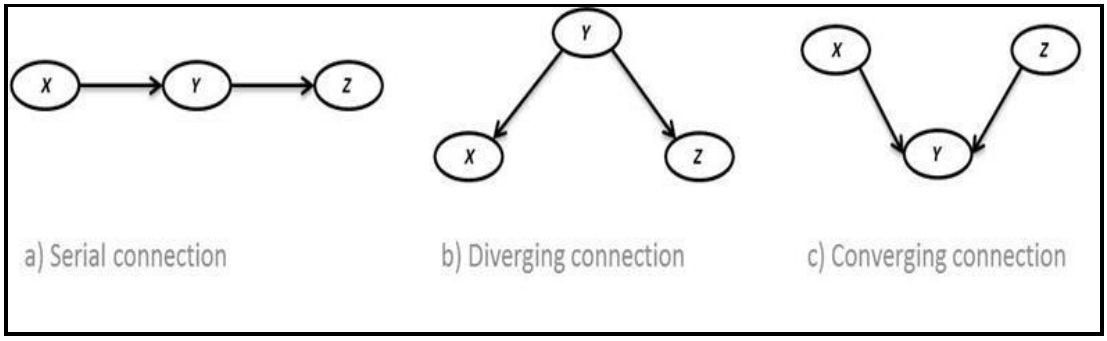


Figure A-2: Types of connection for defining conditional independence in a BN

- Serial Connection: An instance of this connection is illustrated above in *Figure A-2a*. It occurs when a variable or node in a BN is a *parent* of one of the variables and *child* to the other. In any path in a BN with a serial connection, say $\{X \rightarrow Y \rightarrow Z\}$, evidence from *X* through *Z* is blocked iff *Y* is instantiated.

  For all practical BN modeling purposes, if the relationship of the variables in a BN is modeled as a serial connection, for instance $\{X \rightarrow Y \rightarrow Z\}$, but evidence in *X* can influence *Z* after *Y* is instantiated, then in this case, the *serial connection* is not the appropriate representation of the relationship of the variables.

- Diverging Connection: In a BN that is motivated by causal perceptions, the *diverging connection* illustrated above in *Figure A-2b* refers to a situation where two variables have a common cause. In any path in a BN with a diverging connection, say $\{X \leftarrow Y \rightarrow Z\}$, evidence from *X* through *Z* is blocked *iff Y* is instantiated.

- Converging Connection: This type of connection is illustrated above in *Figure A-2c.* In this instance, also referred to as *common cause* connection, probabilistic influence from node *X* can only flow to node *Z iff* the node, *Y*, is instantiated. In other words, nodes *X* and *Z* are *d*-connected given *Y*.

From the foregoing discussion about the structural properties of Bayesian Network, it can be deduced that the stochastic independence assumptions of a BN do have specific dynamic properties which enable the change from independence to dependence (or vice versa), when nodes are instantiated.

A.4 Construction of Bayesian Networks

In the previous sections we provided a brief overview of Bayesian Networks and some related theory. However, not every problem can be adequately modeled by a Bayesian Network; so it is imperative that, given a problem, the analyst/modeler should determine if BN is an appropriate methodology for solving the problem. Kjaerulff & Madsen (2013) state that the use of the BN modeling paradigm is well-suited for problem domains that have some elements of uncertainty associated with the problem definition and scope.

Recent studies indicate that the domain of application of BN is very wide (Pavlenko & Chernyak, 2010; Demirer, Mau, & Shenoy, 2006; Cowell, Verrall, & Yoon, 2007), however, the major consecutive steps involved in constructing a Bayesian Network are:

- *Identify variables of interest and their respective states that define the problem domain*: Variables of a BN (and their respective states) are very important elements of a BN model, and more so, eliciting the structure of a BN manually by identifying relevant variables and states can be an intensive and iterative process. Herein domain expert knowledge and insight with respect to the problem oftentimes determine the level of granularity, and the analyst/modeler will have to make careful trade-off between completeness and simplicity.
- *Define the relationship of variables and network structure:* Probabilistic influence and causality play important roles in Bayesian Networks. In some instances, the directed edges of a BN may either be motivated by either probabilistic influence or causality, or both. However, within the scope of this study, probabilistic influence is a perspective that we have maintained in giving meaning to the directionality of the arcs of BN model proposed in this study. Hence, a directed edge from variable *A* to *B* depicts some form of probabilistic influence between these two variables.
- *Specify the conditional probability for each node:* The relationship between nodes in a BN is quantified by the conditional probabilities, and expressed in terms of a *Conditional Probability Table (CPT)* associated by each node. The joint probability distribution encoded by a BN is a realization of all the conditional probabilities for all nodes in a BN. Research indicates that this step in the construction of a BN, which represent the quantitative part of BN modeling, is

often the most challenging. Though the difficulty and objectivity involved may vary depending on the problem domain.

A.5 Bayesian Network Modeling with *AgenaRisk*®

There are some powerful software tools for building robust Bayesian Networks, such as *Uninet* (a BN software tool for dependence modeling of high dimensional distributions), *BayesiaLab* (a software tool with a suite of capabilities – including machine learning, knowledge modeling, simulation, and optimization – that are based on Bayesian Networks), *Netica* (a software tool for working with Bayesian Networks and Influence Diagrams). However, in this study we have utilized *AgenaRisk*® Bayesian Network software in developing our model. Our choice of *AgenaRisk*® is premised on its sophistication and capacities which are well suited for risk assessment. In contrast to some other BN tools, some of the distinctive features of *AgenaRisk*® which have enabled building the model developed in this study include:

- A powerful algorithm – dynamic discretization algorithm – which enables efficient propagation of influence and building BN models that have both continuous variables and discrete variables (i.e. hybrid model)
- A highly intuitive, easy to use, and sophisticated graphic user interface.
- A wide range of built-in statistical distributions, operators, and expressions (e.g. IF, THEN, ELSE), and some powerful functionalities such as comparative and partitioned expressions for defining Conditional Probability Tables (CPTs). For a *child* node with more than one *parent* node, the Partitioned Expression enables

efficient modeling of the relationship through the use of mathematical expressions that define the relationship of the child node given each parent.