Project Planning: Improved approach incorporating uncertainty

Vahid Khodakarami, Queen Mary University of London
Norman Fenton, Queen Mary University of London and Agena Ltd
Martin Neil, Queen Mary University of London and Agena Ltd

Submitted to Track 15 EURAM2005:

"Reconciling Uncertainty and Responsibility"

Abstract

Project planning inevitably involves uncertainty. The basic input parameters for planning (time, cost and resources for each activity) are not deterministic and are affected by various sources of uncertainty. Moreover, there is a causal relationship between these uncertainty sources and project parameters; this causality is ignored in current state-of-the-art project planning techniques (such as simulation techniques). In this paper we present an approach, using Bayesian network modelling, that addresses both uncertainty and causality in project management. Bayesian networks have been widely used in a range of decision-support applications, but the application to project management is novel. The model we present empowers the traditional Critical Path Method (CPM) to handle uncertainty and also provides explanatory analysis to elicit, represent, and manage different sources of uncertainty in project planning.

Keywords: Project planning; uncertainty; Bayesian networks; CPM.

1 Introduction

Project planning is difficult because it inevitably involves uncertainty. Projects are typically characterised by:

- *uniqueness* (no similar experience)
- *variability* (trade-off between performance measures like time, cost and quality)
- *ambiguity* (lack of clarity, lack of data, lack of structure and bias in estimates)

Yet, although many different techniques and tools have been developed to support better project planning (and these tools are used seriously by a large majority of project managers [Fox 1998, Pollack-Johnson 1998]), quantifying uncertainty is rarely prominent in these approaches.

Since it is still an important open problem, this paper focuses especially on the problem of handling uncertainty in project scheduling. In Section 2 we elaborate on the nature of uncertainty in project scheduling and summarise the current state-of-the-art. Our approach to incorporating uncertainty is to adapt one of the best-used scheduling techniques, *Critical Path Method* (*CPM*) [Kelly 1961], and incorporate it into an explicit uncertainty model (using Bayesian Networks). Thus, in Section 3 we summarise the basic CPM methodology and notation, along with a standard example that we will use in subsequent extensions. In section 4 we present a brief introduction to Bayesian Networks, and describe how the CPM approach can be incorporated. In Section 5 we demonstrate the results of the model and show how it can be used to support project scheduling. Since the research described here represents work in progress, Section 6 suggests the way forward and possible future modifications.

2 The nature of uncertainty in project scheduling

The project management body of knowledge PMBOK identifies *risk management* as a key area:

> ‘It includes the processes concerned with identifying, analyzing, and responding to project risk. It includes maximizing the results of positive events and minimizing the consequences of adverse events’. [Duncan 1996]

Central to risk management is the issue of handling *uncertainty*. [Ward and Chapman 2003] argue that current project risk management processes induce a restricted focus on managing project uncertainty. They believe it is because the term ‘risk’ has become associated with ‘events’ rather than more general sources of significant uncertainty.

In different project management processes there are different aspects of uncertainty. Here we are especially interested in uncertainty in project scheduling. The most obvious area of uncertainty here is in estimating duration for a particular activity. Difficulty in this estimation arises from a lack of knowledge of what is involved rather than from the uncertain consequences of potential threats or opportunities. This uncertainty arise from one or more of the following:

- Uncertainty about the level of available and required resource
- Trade-off between resources and time
- Possible occurrence of uncertain events (i.e. risks)
- Causal factors and inter-dependencies between them
- Lack of previous experience and use of subjective data instead of objective data (expert judgement)
- Uncertainty about the basis of subjective estimation (i.e. Bias in estimation)
- Common casual factors that affect more than one activity (such as organizational issues)

The best-known technique to support project scheduling is the *Critical Path Method* (*CPM*) (which we describe in detail in Section 3). This technique, which is incorporated into the most widely-used project management software tools, is purely deterministic. It makes no attempt to handle or quantify uncertainty. However, another well-used technique *Program Evaluation and Review Technique* (*PERT*) [Miller 1962, Moder 1988] does incorporate uncertainty in a restricted sense, by using a probability distribution for each task. Instead of having a single deterministic value, three different estimates (pessimistic, optimistic and most likely) are approximated. Then the 'critical path' and the start and finish date are calculated by use of distributions' means and applying probability rules. Results in PERT are more realistic than CPM but PERT does not address explicitly any of the sources of uncertainty listed above.

In addition to PERT most recent techniques for handling uncertainty in project scheduling use *Monte Carlo simulation* [Cook 2001]. For example, *Pertmaster*

[Pertmaster 2005] has incorporated Monte Carlo simulation to provide project risk analysis in time and cost. However, the Monte Carlo approach has attracted some criticism. For example, [Williams 2004] describes how Monte Carlo simulation results can mislead project managers by ignoring potential management actions to bring late running project back on track. [Van Dorp and Duffey 1999] explain another weakness of Monte Carlo simulation, namely the assumption of statistical independence of activity duration in a project network. Of more concern to us is the issue that Monte Carlo simulation tools like Pertmaster (being event-oriented) do not identify the source of uncertainty. As argued in [Ward and Chapman 2003] managing uncertainty in project planning is not just about managing perceived threats, opportunities and their implication. A proper uncertainty management provides identifying various sources of uncertainty, understanding the origins of them and then managing them to deal with desirable or undesirable implications. Our challenge is to address these specific issues, and we do so by incorporating the CPM approach into the most widely used formalism for modelling causality and uncertainty, Bayesian Nets. Next we provide an overview of the CPM methodology and notation.

3 CPM methodology and notation

CPM [Moder 1988] is a deterministic technique which, by use of a network of dependencies between tasks and given deterministic values for task durations, calculates the earliest network (the 'critical path') which is the earliest time for project completion.

CPM has the following steps:

- Identify the individual activities (tasks). This usually is done by work breakdown structure. Based on size and complexity of the project, the project manager specifies the list of all required activities during the project.
- Determine the dependency between activities. This requires listing immediate predecessors, which must complete before each activity can start.
- Draw project network (diagram). This shows all activities and their dependencies.
- Estimate activities durations. This is usually done by historic data or expert judgement. For each activity a single point duration is estimated.
- Identify the critical path. This is the longest path of the network and the earliest time for project completion.
- Update the CPM network. As the project progresses and more information about activities is available, the network structure and calculation are updated.

The critical path can be identified by determining the following parameters for each activity:

D - Duration

ES - earliest start time

EF - earliest finish time

LF - latest finish time

LS - latest start time

The earliest start and finish times of each activity are determined by working forward through the network and determining the earliest time at which an activity can start and finish considering its predecessor activities. For each activity j:

ES_j = Max [$ES_i + D_i$ | i one of the predecessor activities]

$EF_j = ES_j + D_j$

The latest start and finish times are the latest times that an activity can start and finish without delaying the project and are found by working backward through the network. For each activity i:

LF_i = Min [$LF_j – D_j$ | j one of the successor activities]

$LS_i = LF_i – D_i$

The activity's slack (which is the amount that the activity’s duration can be increased without increasing the overall project completion time) is the difference in the latest and earliest finish of each activity. A critical activity is one with no slack time and should receive special attention (delay in a critical activity will delay the whole project). The critical path then is the path(s) through the network whose activities’ have minimal slack.

Example (1): Consider a small project with five activities A, B, C, D and E. Activity A is processor of B and C, and activities C and D are predecessors of E. Figure 1 shows a simple project network.

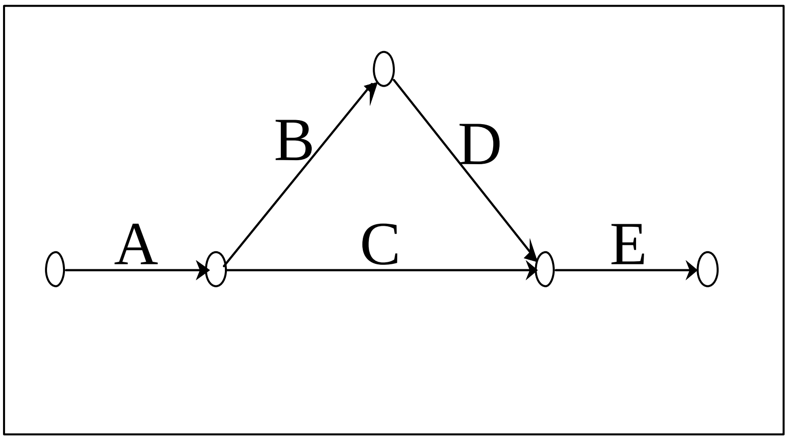


Figure 1 project network for example (1)

Table (1) summarizes the results. Activities A, C and E with no slack time are critical and the overall the project takes 20 days.

Table 1 Activities' time (days) for example (1)

Activity	Duration	ES	EF	LS	LF	Slack
A	5	0	5	0	5	0
B	4	5	9	9	13	4
C	10	5	15	5	15	0
D	2	9	11	13	15	4
E	5	15	20	15	20	0

The CPM approach is very simplistic and provides very useful information about a project and its activities' schedule. However, because of its' single point estimate assumption it is not useful for realistic projects. The challenge is to incorporate the inevitable uncertainty.

4 Proposed BN solution

Bayesian Networks (BNs) are recognised as a mature formalism for handling causality and uncertainty [Heckerman et al 1995]. In this section we provide a brief overview of BNs and describe a new approach for scheduling project activities in which. CPM parameters (i.e. ES, EF, LS and LF) are determined in a BN.

4.1 Bayesian Networks: An overview

Bayesian Networks (also known as Belief Networks, Causal Probabilistic Networks, Causal Nets, Graphical Probability Networks, Probabilistic Cause-Effect Models, and Probabilistic Influence Diagrams) provide decision-support for a wide range of problems involving uncertainty and probabilistic reasoning. Example real-world applications can be found in [Heckerman et al 1995, Fenton et al 2002, Neil et al 2001].

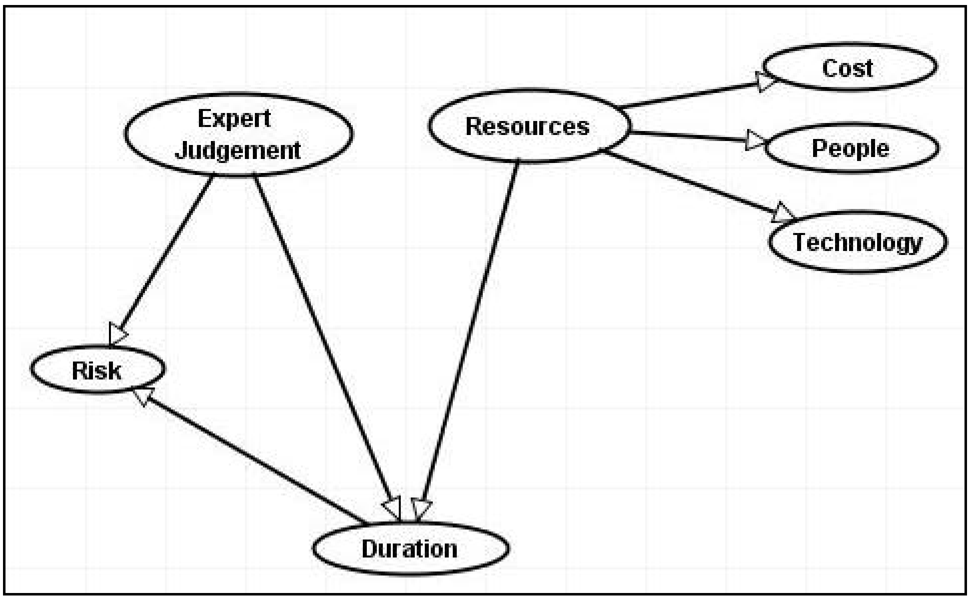


Figure 2 Bayesian Network for Activity Duration

A BN is a directed graph (such as the one shown in Figure 2), together with an associated set of probability tables. The graph consists of nodes and arcs. The nodes represent uncertain variables, which may or may not be observable. The arcs represent causal or influential relationships between variables.

The main use of BNs is in situations that require statistical inference. In addition to statements about the probabilities of events, the user knows some *evidence*, that is some events that have actually been observed, and wishes to infer the probabilities of other data, which have not as yet been observed. These observed values (evidence) represent a posterior probability, and by applying Bayes rule in each affected node, they can influence other BN nodes via propagation, modifying the probability distributions. There are numerous commercial tools that enable users to build BN models and run the propagation calculations. With such tools it is possible to perform fast propagation in large BNs (with hundreds of nodes). In this work we have used [AgenaRisk 2005], which is especially well-suited to this kind of application.

In summary, BNs have the following advantages that make them highly suitable for the project planning domain:

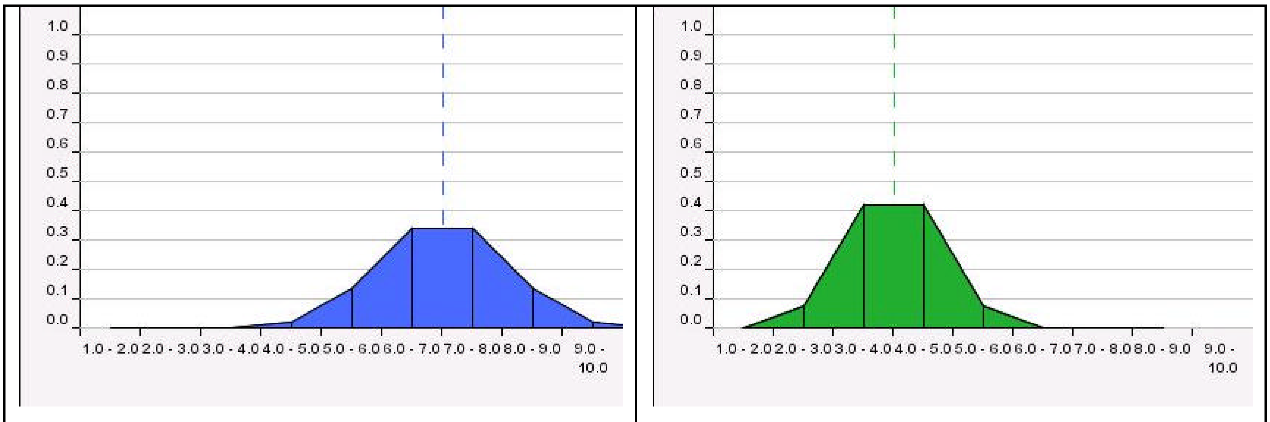
- Explicitly quantify uncertainty
- Reason from effect to cause as well as from cause to effect (propagation is both 'forward' and 'backward')
- Overturn previous beliefs in the light of new data ('explaining away')
- Make predictions with incomplete data
- Combine subjective and objective data
- Arrive at decisions that are based on visible auditable reasoning

BNs, as a tool for decision support, have been deployed in domains ranging from medicine to politics. We believe BNs potentially address many of the 'uncertainty' raised in Section 2. In particular, by incorporating CPM-style scheduling into a BN framework we can properly handle uncertainty in project scheduling.

4.2 BN for Activity Duration

Figure 2 above shows a prototype BN that we have built to model uncertainty sources and their affects on duration of a particular activity. The model contains variables that capture the uncertain nature of activity duration. 'Expert Judgement' is the first estimation of activity duration; it is estimated based on historic data or previous experience. 'Resources' is any affecting factor, which can increase or decrease the activity duration. It is a ranked node, which for simplicity here we restrict to three levels: low, average and high. The level of resources can be inferred from so called 'indicator' nodes. Hence, the causal link is from the 'resources' to directly observable indicator values like the 'cost' and the level of available 'people' (in terms of quantity and quality) and the level of available 'technology'. There are many alternative indicators. An important and novel aspect of this approach is to allow the model to be adapted to use whichever indicators are available.

The power of this model is better understood by showing the results of running it under various scenarios in the AgenaRisk software (all subsequent figures are outputs from the tool). We can enter observations anywhere in the model to perform not just predictions but also many types of trade-off and explanatory analysis. So, for example, we can enter observations for 'expert judgement' and 'resources' and let the model show us the distributions for ‘duration’. Figure 3 shows how the distribution of an activity changes when the level of its available resources changes.



a) Resource is ‘Low’ b) Resource is ‘High’

Figure 3 Probability distribution for ‘Duration’ (number of days) changes when the level of ‘Resources’ changes

Another possible analysis in this model is the trade-off analysis between ‘duration’ and ‘Resources’. When there is a time constraint for activity duration and we are interested to know about the level of required resource. For example, Figure 4, shows the resulting distribution for the node ‘Resources’ if the first estimation for an activity is five days but it needed to be finished in three days; note that the required resource is most likely to be ‘high’.

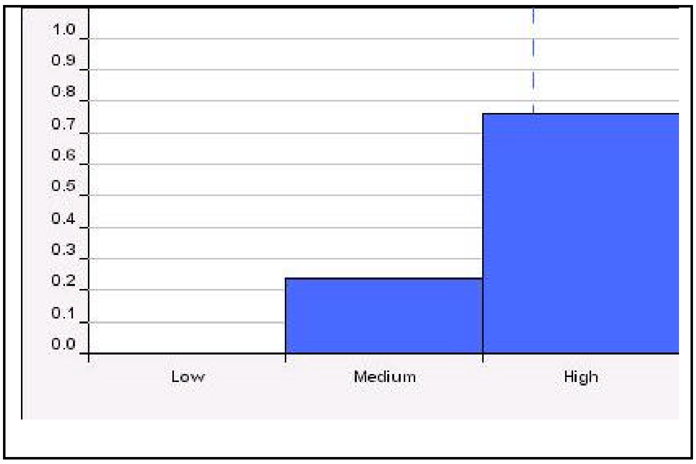


Figure 4 Level of required ‘Resources’ when there is constraint on ‘Duration’

The model also provides additional analytical information about the posterior probability of 'risk' and 'resources'. When actual duration of an activity is known as the project progresses, by comparing 'actual' and 'predicted' durations, the model can update the distribution for 'risk' and 'resources'. In turn this distribution can be used for later analysis or for other activities with common casual dependencies. Figure 5 shows the distribution of 'resources' for an activity that was originally estimated at five days but actually lasted seven days.

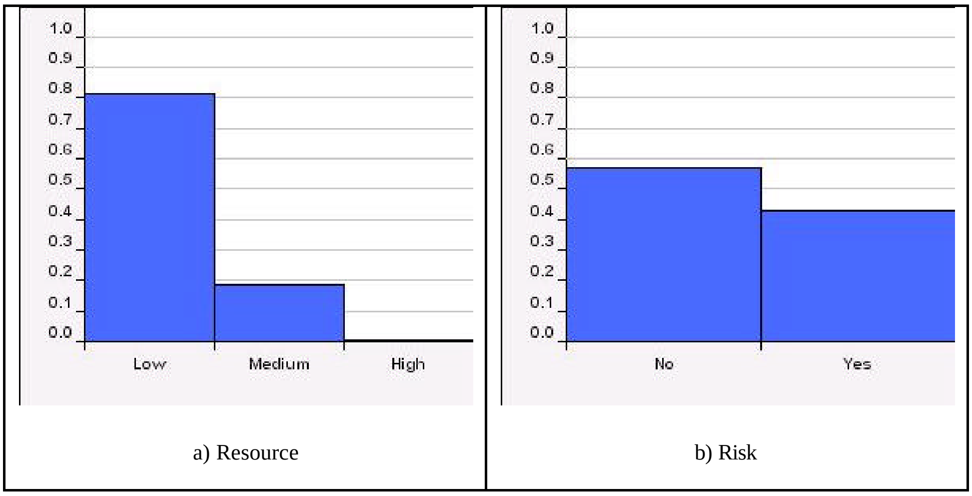


Figure 5 Posterior probability for a) Resource and b)Risk when the actual 'Duration' was 7

In the next section we use this model to reason about more uncertain activities.

4.3 Mapping CPM to BN

As we have seen, the main component of CPM networks is *activity*. Activities are linked together to represent dependencies. In order to map a CPM network to a BN we first need to map a single activity. We take the activity parameters identified in Section 3 and represent each as a variable (node) in the BN:

Duration (D): This is the most uncertain variable and its values are inferred from the risk factors and common causal dependencies. It may have various probability distributions. For uncertain activities which need more analysis, the duration network (explained in section 4.2) can be used.

Earliest Start (ES): This gets its value from the predecessor activity/activities. For the first activity of the project ES = 0 and for other activities ES is the maximum of the predecessor activity's EF.

Earliest Finish (EF): This is a derived node, which adds the ES and D.

Latest Finish (LF): This gets its value from the successor activity/activities. For the last activity of the project LF=EF and for other activities LF is the minimum of the successor activities' LS.

Latest Start (LS): This is a derived node, which subtracts D from LF.

Figure (6) shows a schematic model of the BN fragment associated with an activity. It clearly shows the relation between the activity parameters and also the relation with predecessor and successor activities.

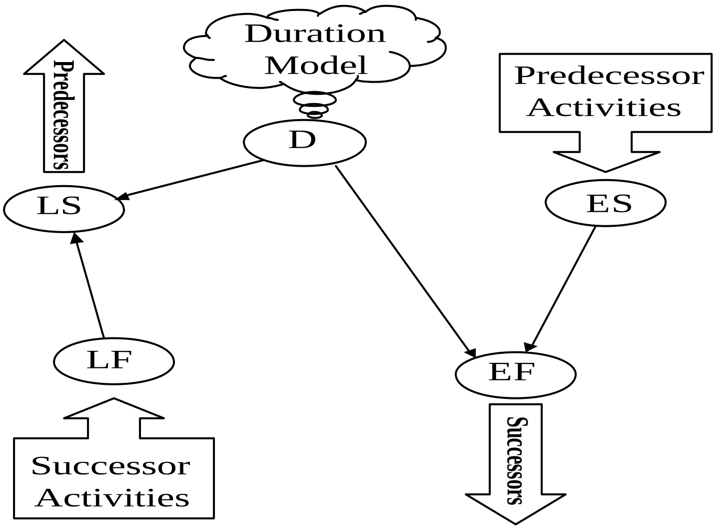


Figure 6 Schematic of BN for an activity

The next step is to define the connecting link between dependent activities. The forward pass method in CPM is mapped as a link between EF of predecessor activity to ES of successor activity. The backward pass method in CPM is mapped as a link between LS of successor activity to LF of predecessor activity.

Figure 7 shows the BN representation of example (1). Every activity has 5 associated nodes. Forward pass calculation of CPM is done through connection between ES and EF. Activity 'A', the first activity of the project, has no predecessor, so its ES is set to zero. 'A' is predecessor for 'B' and 'C' so EF of 'A' is linked to ES of 'B' and 'C'. EF of 'B'is linked to ES of its successor, 'D'. And finally EF of 'C' and 'D' are connected to ES of 'E'. In fact ES of 'E' is the maximum of EF of 'C' and 'D'. EF of 'E' is the earliest time for project completion time.

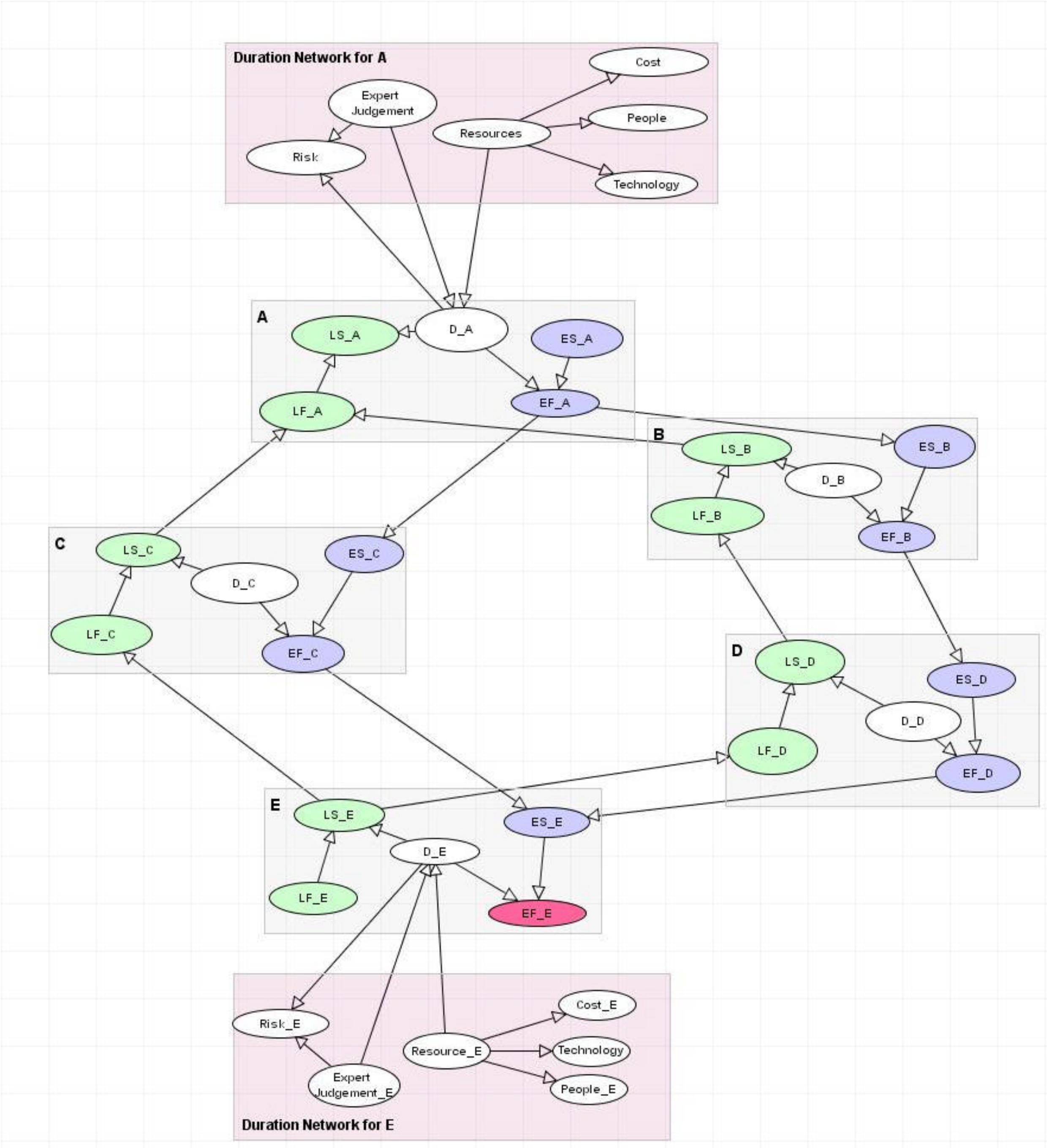


Figure 7 Overview of BN for example (1)

The same approach is used for backward CPM calculation with connecting LF and LS. Activity 'E' that is the last activity of project has no successor, so its LF is set to EF. 'E' is successor of 'C' and 'D' so LS of 'E' is linked to LF of 'C' and 'D'. LS of 'D' is linked to LF of its predecessor 'B'. And finally LS of 'B' and 'C' are linked to LF of 'A'. LF of 'A' is the minimum of LS of 'B' and 'C'.

For simplicity in this example, we suppose that activities 'A' and 'E' are more risky and need more detailed analysis. For other activities the uncertainty about 'duration' is expressed simply by a normal distribution.

4.4 Object Oriented approach

It is clear from Figure 7 that even simple CPM networks lead to fairly large BNs. In fact this complexity can be handled using the Object Oriented Bayesian Network (OOBN) approach [Koller and Pfeffer 1997]. This approach supports a natural framework for abstraction and refinement, which allows complex domains to be described in terms of inter-related objects. In addition to generalization, OOBN supports hierarchy and inheritance.

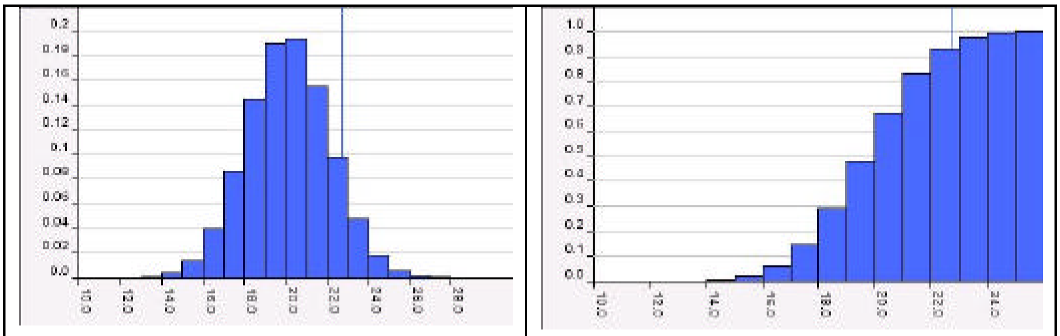
In our model each activity, as the basic unit of a network, is an object and the internal parts of the activity subnet (Figure 2) are encapsulated within the object. The OOBN approach can also significantly improve the performance of inference in the model. Although the OOBN approach to this particular problem is beyond the scope of this paper, the key point to note is that there is an existing mechanism (and implementation of it) that enables us to genuinely 'scale-up' the proposed solution to real-world projects.

5 Results

In this section we explore scenarios of the BN model derived from the simple CPM network of example 1.

The main objective is to predict project completion time (i.e. the earliest finish of E, the last activity of the project) in such a way that it fully characterises our uncertainty.

Suppose the initial estimation of activities' duration is the same as Table 1. Suppose the resource level for 'A' and 'E' is 'medium'. If the earliest start of 'A' is set to zero, the distribution for project completion is shown in Figure 8a. The distribution's mean is 20 days as was expected from CPM analysis. However, unlike CPM our prediction is not a single point and its variance is 4. Figure 8b illustrates the cumulative distribution of finishing time, which shows the probability of completing the project before a given time. For example, with probability of 90% the project will finish in 22 days.

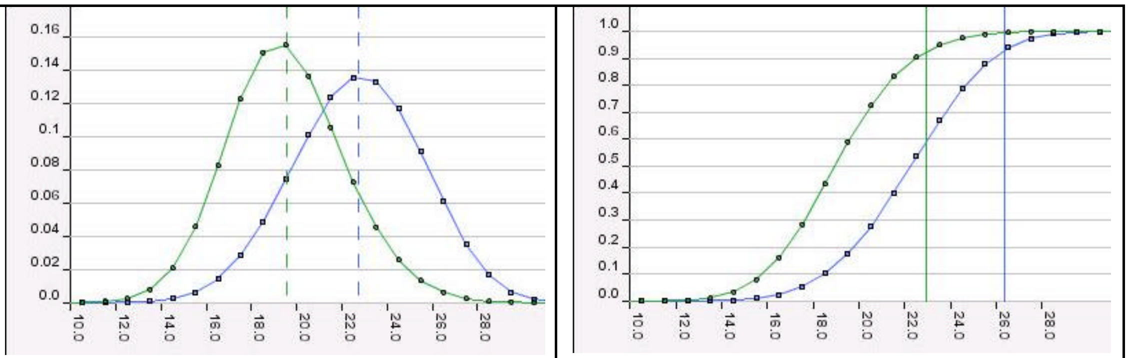


a) Probability Distribution

b) Cumulative Distribution

Figure 8 Distribution of project completion (days) for main Scenario in example (1)

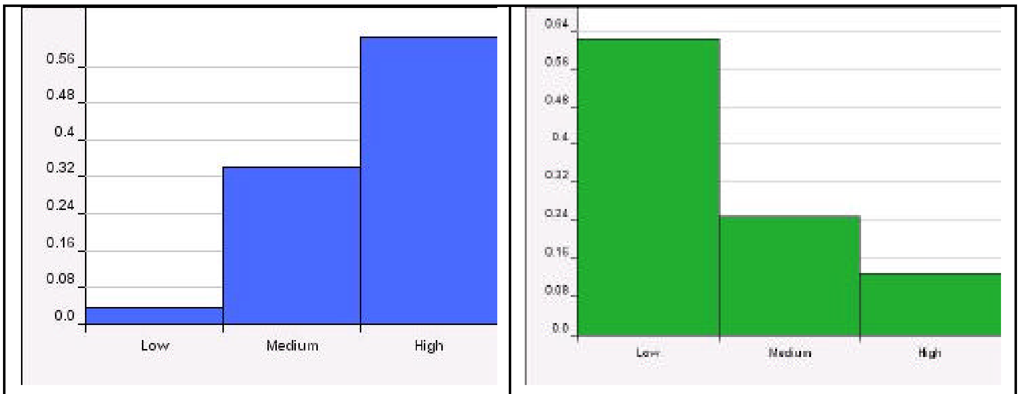
In addition to this baseline scenario, by entering various evidence (observations) to the model we are able to analyses the project plan from different aspects. For example, one scenario is to see how changing the resource level affects the project completion time. Figure 9 compares the distributions for project completion time when level of people is ‘low’ against ‘high’. When the level of ‘people’ changes from ‘low’ to ‘high’ the mean of finishing time changes from 22.7 days to 19.5 days and the 90% confidence interval changes from 26.3 days to 21.2 days.



a) Probability b) Cumulative

Figure 9 Change in project time distribution (days) when level of ‘People’ changes from ‘low’ to ‘high’

Another useful analysis is when there is a constraint on project completion time and we want to know how much resource we need. Figure 10 illustrates this trade-off between project time and required resources. If the project needs to be completed in 18 days (instead of baseline 20 days) then the resource required for ‘A’ most likely must be ‘high’; if the project completion is set to 22, the resource level for A changes moves significantly in the direction of ‘low’.

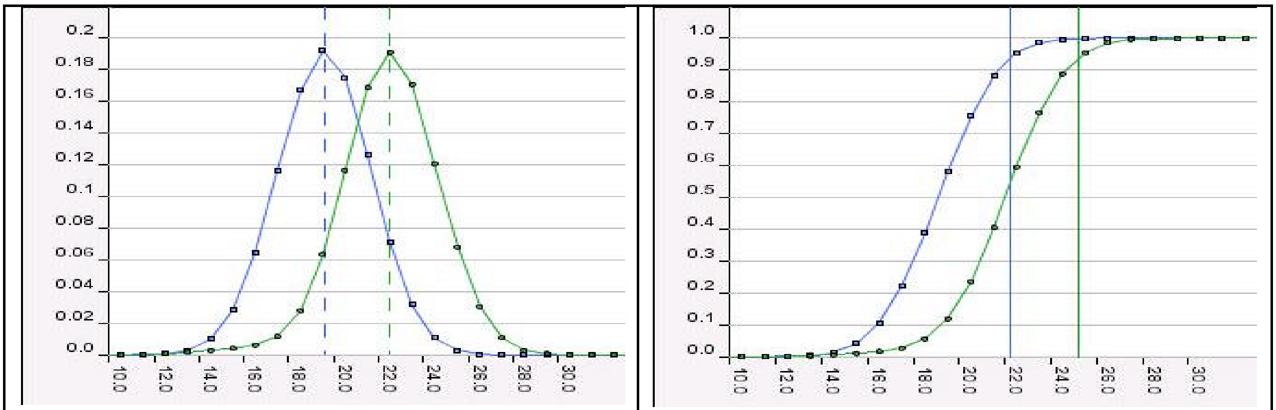


a) Finish time=18 days

b) Finish time=22 days

Figure 10 Trade-off between project time and required resources

The next scenario is the investigation of the impact of risk in activity A on the project completion time. Figure 11 shows this scenario. When there is a risk in activity A the mean of the distribution for project completion time changes from 19.5 to 22.4

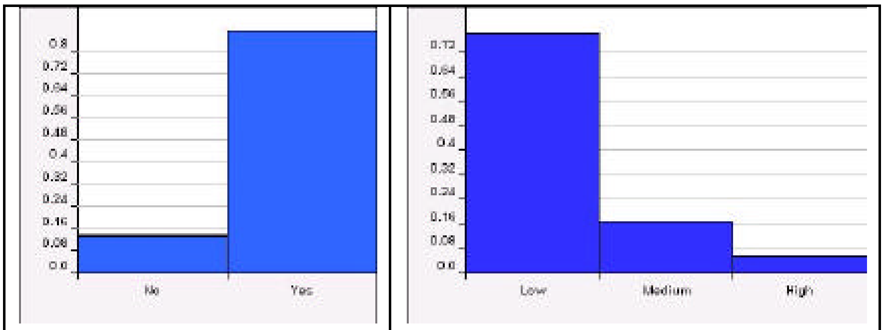


a) Probability

b) Cumulative

Figure 11 The impact of risk in activity A on project completion time (days)

Another powerful advantage of BNs is their potential for parameter learning, which is shown in the next scenario. Imagine activity A *actually* finishes in 7 days even though it was originally estimated as 5 days. By entering this observation in EJ_A and EF_A the model gives the probability of resource and risk for activity A as it is illustrated in Figure 12.

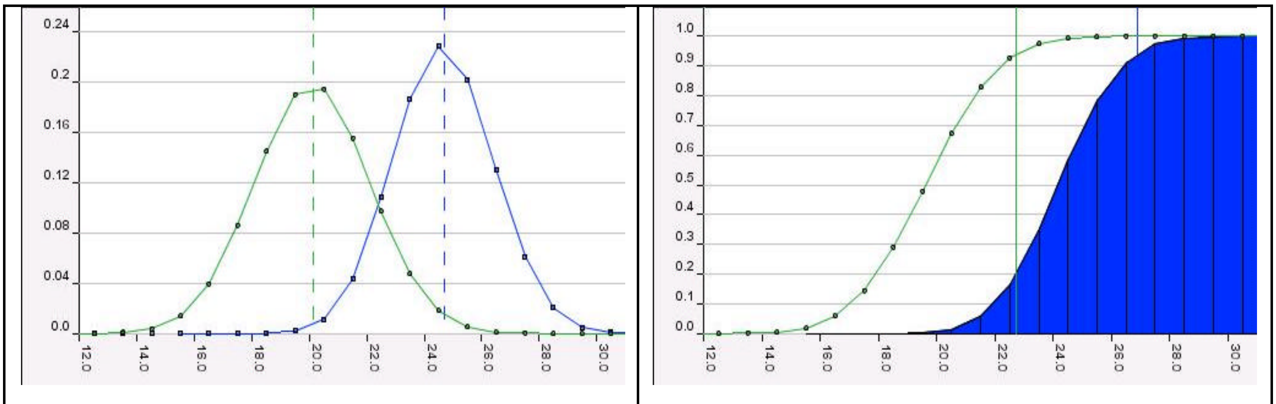


a) Risk

b) Resources

Figure 12 Probability for A's risk and resource when the actual duration is 7 days rather than initially estimated 5 days

If activities have common causal risks, the learnt probability for A's risk inform activity E. Figure 13 compares the distributions of completion time when learned information from risk of activity A is entered to risk of activity E.



a) Probability

b) Cumulative

Figure 13 Completion time (days) based on learned parameters compare with baseline scenario

Another application of parameter learning in these models is the ability to incorporate and learn about estimation bias. So if we get several observations in which actual task completion times are underestimated the model learns that this may be due to bias rather than unforeseen risks and this information will inform subsequent predictions.

6 Conclusions and way forward

Handling risk and uncertainty is increasingly seen as a crucial component of project management and planning. Researchers (and some practitioners) have thus moved on from the generation of MS Project type tools that treat project planning and scheduling purely deterministically. However, most current techniques for handling risk and uncertainty in project planning (i.e. simulation based techniques) are often event-oriented and try to model the impact of possible 'threats' on project performance. They ignore the source of uncertainty and the causal relations between project planning parameters.

We have proposed a new approach that enables us to incorporate risk, uncertainty and causality. Specifically, we have shown how a Bayesian Network model can be generated from a project's CPM network. Part of this process is automatic and part involves identifying specific risks (which may be common to many activities) and resource indicators. The benefit of incorporating the project schedule into a BN is that we can then bring the full weight and power of BN analysis to bear on the problem of project scheduling. This means we can

- Make explicit different sources of uncertainty and use this to inform project planning.
- Express uncertainty about completion time for each activity and the whole project with full probability distributions
- Model the 'trade-off' between 'time' and 'resources' in project activities
- Use 'what-if?' analysis for finding the level of required resources given constraints like, for example, a specific completion time

- Learn from data (as the project progresses and more information become available) so that predictions become more relevant and accurate

The application of the approach was explained by use of an artificial simple example. In order to scale this up to real projects with many activities the approach must be extended to use the so-called Object Oriented BNs.

The BN approach could be extended and improved by:

- Including additional uncertainty sources in the duration network
- Handling dynamic parameter learning as more information becomes available when project progresses
- Handling common causal risks which affect more than one activity
- Handling management action when the project is behind its plan

Hence we feel that the BN approach provides a potentially revolutionary way forward for tackling uncertainty in project planning.

7 References

AgenaRisk, www.agenarisk.com, 2005

Cook M. S, “Real-world Monte Carlo Analysis”, Proceeding of PMI Annual Seminars and Symposium, Nashville, 2001.

Duncan W R."A Guide To the Project Management Body of Knowledge", Project Management Institute standard committee, 1996.

Fenton NE, Krause P, Neil M, "Software Measurement: Uncertainty and Causal Modelling", IEEE Software 10(4), 116-122, 2002.

Fox T L, Spence J W, "Tools of the trade: A survey of project management tools", Project Management Journal, Vol. 29, 20-28, Sep98.

Heckerman D, Mamdani A, Wellman M, "Real-world applications of Bayesian networks", Comm ACM, 38(3), 25-26, 1995.

Kelly JE, "Critical path planning and scheduling mathematical bases", Operations Research; Vol. 9, 246-320, 1961.

Koller D, Pfeffer A, "Object-Oriented Bayesian Networks", In D. Geiger and P. Shenoy, editors, Proceeding of the Thirteenth annual Confrerence on Uncertainty in Artificial Intelligence (UAI-97), San Francisco, California, 302-313. Morgan Kaufman Publishers, August 1997.

Miller R. W., "How to plan and control with PERT', Harvard Business Review, 93-104.

Moder J, "Network techniques in project management", Project Management Handbook, 2nd Edition, New York: Van Nostrand Reinhold, 324-373, 1988.

Neil M, Fenton N, Forey S and Harris R, “Using Bayesian Belief Networks to Predict the Reliability of Military Vehicles”, IEE Computing and Control Engineering J 12(1), 11-20, 2001.

Pertmaster Project Risk Software, www.pertmaster.com/products, 2005

Pollack-Johnson B, “Project management software usage patterns and suggested research directions for future development”, Project Management Journal, Vol. 29, 19-29, Jun98.

Van Dorp J.R, Duffey M.R., “Modelling Statistical Dependence in Risk Analysis for Project Networks”, International Journal of Production Economics, Vol. 58, 17-29, 1999.

Ward S, Chapman C, “Transforming project risk management into project uncertainty management”, International Journal of Project Management; Vol. 21, 97-105, 2003.

Williams T, “Why Monte Carlo Simulation of Project Network Can Mislead”, Project Management Journal, Vol 35(3), 53-61, Sep 2004.