A Probabilistic Approach for Project Cost Estimation Using Bayesian Networks

Hossein Raoofpanah¹, Khadije Hassanlou²

¹ Raja University Ghazvin, Iran. Email: h.raoofpanah@yahoo.com
² Industrial Engineering Department, Raja University, Ghazvin Iran. Email: kh_hassanlou@iust.ac.ir

Abstract: Risk measurement is one of the main stages in Project Risk Management. It quantifies risks and assesses their impact on project’s outcomes (time, cost and quality). Monte Carlo simulation, as the best practice, has been used to developed several models to analyze and quantify risks in projects. Bayesian Networks (BNs), as a powerful technique for decision support under uncertainty, have attracted a lot of attention in different fields. This paper aims to use BN capabilities to introduce a new approach for project cost risk modeling. The new approach explicitly quantifies uncertainty in project cost and also provides an appropriate method for modeling complex relationships and factors in projects such as: causal relation between variables, common causal factors, formal use of expert judgments, and learning from data to update previous beliefs and probabilities.


Keywords: Project Risk Analysis; Cost Estimation; Bayesian Networks; Simulation

1. Introduction

Cost is one of the main indexes of decision making in different level of project management. In today’s competitive world, cost management is in addition of preservation levels of the quality and performance, it plays a main role. So assessment and cost control attract attention, significantly and based on what many experts said the weak estimating of cost is the main factor in the failure of project. Often, the real cost of projects is beyond what is estimated (In the sources (1-5) brings samples and reports from increasing of real project cost), this matter exist even in occasion that estimating in details and in the total level of project accomplished by experienced group. When, we agree to undertake 10,000$ project in amount of 70000 $, before starting every work in project, we are subjected to cost overrun, schedule slippage and deficient performance. Thus, always, cost has been proposed as a criterion of successful project. Because of uncertainty nature of cost, some part of the risk management allocates to risk analysis. Risk analysis of cost prepare more definitive, realistic and correct from project costs. Estimating of cost is a dynamic process and depends on which stages we are, level of estimating accomplishes in different forms and for each especial aim and purpose accomplish. In the source of these levels, estimating bring in various names and forms [7-9]. But totally we can consider 3 levels of conceptual estimate, preliminary estimate and definitive estimate. Traditional methods of estimating of cost find in literature project management and cost management [8, 10]. This method achieves point estimate and thus due to the lack of cost risk, isn’t real. In opposite, there are quantitative risk analysis methods in following identification and modeling of uncertainties exist in cost elements. The main concentration of this paper is on the definitive estimate. In this stage, the total details of work define very well such as worked breakdown structure (WBS). These estimates are the most definitive estimates and estimate is done in bottom –up. Now, One of the main techniques using for this kind of estimate, is Monte Carlo simulation (source of 1,7,11 and 12 describe the process). This paper tries with using one of the techniques of Artificial Intelligence presents the better analysis of cost risk in definitive estimate or bottom- up in addition of benefiting from approach of Monte Carlo Simulation. In addition, already used from methods of artificial intelligence in estimating of cost. Gunaydinet al. (2004) and Zhang et al (1997) , in their studies uses approach of neural network for preliminary estimate of cost in building project and packing product, respectively[14,13]. Khodakarami (2007) and Luu (2009) present modeling from Bayesian Networks in the topic of discussion of schedule risk [15, 16] but, by now doesn’t benefit from ability of Bayesian Networks for management of uncertainty in estimate of cost. Here, we consider this kind of modeling of risk in estimating the project cost or in the subdivision of project. The weakness of previous methods is that they used judgment and human information in form of systematic. Because of deficit objective information especially in new and difficult situation such as some IT projects, it seems essential using artificial methods that benefits from judgment and human information. Bayesian Networks having ability of accepting additional and mental information. This approach in situations that doesn’t have any rich historical information has ability of updating information. Moreover, in confronting with
problem of cost risk analysis Bayesian Networks can better modeling the cause and effect relation between cost element and with definition of different scenario for one or set of elements of cost infer others costs elements. So, with adding secondary network could model the effect of risks of real world in projects cost. This paper tries to introduce advanced abilities of Bayesian Networks in better modeling and as a result in more effective project cost management. The structure of the following paper is as follows. In the next section describes definitive estimate of cost project with using Monte Carlo simulation as the superior approach in the quantitative cost risk analysis. In the third section introduce Bayesian Networks. In the fourth section proceeds the structure and logic of presented model. With helping numerical example describes the new abilities of models in cost risk analysis. And finally, in fifth section presents conclusion and proposing for future studies.

2. Quantitative cost risk analysis with simulation

After recognizing details of project until last level and accomplishing schedule for activities, definitive estimate are able to pursue and do. As above mentioned the main approach in this level of estimate is using the Bottom–up method, namely, the components of cost combine with each other in the lowest level of WSB and conclude the major costs. For exerting the uncertainties existing in costs items, recognizing probability distribution for each component of cost and combine with each other in an appropriate methods that usually is simulation.1 In order to achieve an appropriate probability distribution relating to cost items, the main approach is in this arrangement that firstly triplet estimate pessimistic, optimistic and the most likely from the two main source of information conform to them namely past data and in the view of expert of extraction and is a continuous probability of distribution that usually is triangular distribution (fig.1).2 In this methods in order to neutralize the effect of deflection resulting from optimistic, the amount of highest and lowest consider around 90 percent of density of probability. For instance, consider figure 1 that is a triangular distribution for amounts of 300, 400 and 800. Note that probability of 10 percent in equality doesn’t put in both ends of distribution. It puts 2 percent in the left heads and in right side 8 percent. This skewness distinct with considering to the place of the probabilistic amount in proportion to the highest and lowest amount. Other approaches exit expect above approach that in view of expert benefit and doesn’t need to formal probability distribution. Source [12] is a review on existence texts in the fields of extraction of mental probabilities in cost risk analysis.

When distributions combine, determination of affiliation or independence between them is very important. Lack of considering of this affiliation causes underestimate. For example, the weak election of supplier causes the decrease of productivity in some components of project and increases their cost, or if the risk creates in two part, the twin effect of them might be noticeably more than effect of each of them in form of single (all kind of affiliation describe in source [7]). Lack of recognition of positive correlation among the cost element causes decrease of estimate of risk. In the end, they combine with each other with using tools of simulation of distribution and outcome achieve as form of figure 2. The average value can achieves as the most probable cost and probability of the increasing of cost project from a definite amount (more details and new works have been done for developing of this method will find in sources [7], [10],[18]).

1 Instead of simulation numerical integrated process such as distance approach and controlled memory Chapman and moment method can apply for combining the distribution. These methods need special multipurpose software and simplified hypothesizes. So this matter decrease the capability for being practical [7].

2 In some educational texts suggested that take percentile from experts and then conform one normal distribution , leg normal o beta to these value.
addition, determination of coefficient of correlation will be difficult for expert in dealing with many working probability distribution. The other case is that this mechanism doesn’t have the characteristic of learning. In this method we observe and analysis the changes of response with changing of income variable with using the analysis of sensitive but, we can’t infer distribution of income variable with definition of different scenario for total cost (or parts of cost). The proposed model in this paper removes this decrease, in addition of having new abilities.

3. Bayesian Networks

BNs (also known as Belief Networks, Causal Probabilistic Networks, Graphical Probabilistic Networks and Probabilistic Cause-effect Networks) are powerful tools for knowledge representation and reasoning under uncertainty. A BN consists of a set of nodes (representing variables) and a set of directed edges (representing causal influences between variables) between variables. Each variable has a finite set of mutually exclusive states. The variables together with the edges form a directed acyclic graph (DAG) (a directed graph is acyclic if there is no directed path \( A_1 \rightarrow \ldots \rightarrow A_n \) such that \( A_1 = A_n \)). To each variable ‘A’ with parents \( B_1, \ldots, B_n \), a conditional probability table \( P(A|B_1, \ldots, B_n) \) is assigned. If the variable has no parents then the table reduces to the unconditional probabilities \( P(A) \) (i.e. prior probability). One important property of BNs is their ability to represent the joint probability distribution \( P(A_1, \ldots, A_n) \) for all the variables \( A_1, \ldots, A_n \) in a compact form. This is done by use of the ‘chain rule’, which says in a BN the full joint probability distribution is the product of all conditional probabilities specified in the BN \( c \):

\[
P(A_1, \ldots, A_n) = \prod_{i} P(A_i|A_1, \ldots, A_{i-1})
\]  

(1)

The more compact representation of the joint probability makes the probability calculation easier. If we have access to the joint probability distribution, then we can calculate the marginal probability for any variable, \( P(A_i) \), and also the conditional probability of \( P(A_j|A_i=a_i) \). BNs address the problems of storing and representing the joint probability distribution of a large number of random variables and also doing Bayesian inference with these variables. The predictive capability of a BN enables us to infer from cause to effect (from parent to child). The diagnostic capability of the model enables us to infer from effect to cause (from child to parent). During development of a BN, variables (nodes) can be easily added or modified. The graphical nature of BNs allows variables to be conveniently added or removed without significantly affecting the remainder of the network. The advantages of BNs are now widely recognized and they are being successfully applied in diverse fields [20]. During the last decade, researchers have incorporated BN techniques into easy-to-use toolsets, which in turn have enabled the development of decision support systems in a diverse set of application domains.

As the number of network nodes and edges increase, the volume of calculation of combined distribution of the nodes increases exponentially. Developing an algorithm for calculating conditional probabilities and inference in a BN is an NP-hard problem. However, several accurate (for specific networks) and approximate algorithms have been presented and several software tools for developing BN models are available. These tools provide a graphical editor for building the BN and also a runtime module, which takes care of probabilistic calculation and evidence transmission. With such tools it is possible to build a BN and also perform the propagation algorithm in a reasonable amount of time. In this paper, AgenaRisk software [21] has been used, because in addition to discrete random variables, it can also model continuous variables.

4. Proposed model

As already mentioned, definitive estimate of cost achieve by using the last level of WBS. In this section we present a model in which the components of the last level of WBS and the cost of delivery items of project mapping to Bayesian Networks. This part includes some subdivision; in the first section describes the structure of model. Second section allocate to a numerical example from proposed model. In the third, fourth and fifth sections have shown the new abilities of the model in project cost risk analysis including more possibility of considering the relation of affiliation, balance of cost with different parameters of project, and ability of discovering the cause of cost deflection. In the end of this section will discuss about the method of assessment of model.

4-1. Structure of model

Structure of methodology of this model has shown in form of schematic figure 3.

![Figure 3. Structure of methodology of model BN for estimating cost of project.](http://www.lifesciencesite.com)
The process of supplying this method come as following:

Corresponding to approach of bottom to up in estimating of cost, main income is necessary for WBS. Supplying WBS of project and perfect detail of work including package of working and item of delivery is necessary. It estimates the Activity on Node (AON) relating to last level of WBS supplying and duration of time of every activity. It distinguishes Milestone in respect of budgetary and cost control (e.g. monthly or none-uniform milestone). It draws the Gantt chart relating to activity on node of project with this milestone. In lieu of every element of cost (Activity Node Networks), a node make in Bayesian Networks. This rectangular node relates from the left side to right side with each other. Circle nodes of NPT consist of triangular distribution that according to mentioned approach in section 2 extract from judgment information. Rectangular nodes of NPT consist of sum of their parents nodes. In this way the whole cost elements of one project combine base on their relating schedule. The last rectangular node probability distribution project cost shows in the last periods or total project cost. Making rectangular nodes that is proportional milestone and milestone, helping us in addition estimate of the final project cost, estimating cost in the end of special period be available. This structure make possible the more concentration on each period, limitation of budgetary of that period, balance of cost with criteria of time and quality and cost control during project. In the next section, it describes with using an example with above abilities.

4-2. Examples

Figure 4 shows the structure of WBS of a software project and node networks activities of last level. Table 1 contains information of schedule of activities and distribution cost of every activity. Suppose mentioned milestone in algorithm (steps 3) equality one month. Figure 4 shows naming nodes, their relations and outcome of preliminary model.

![Figure 4. a) WBS and nodes network. b) Activity of project of example.](image1)

<table>
<thead>
<tr>
<th>Activity</th>
<th>Duration of Time (Day)</th>
<th>Distribution Cost (dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>20</td>
<td>(5000,500000,600000)</td>
</tr>
<tr>
<td>a2</td>
<td>30</td>
<td>(5000,500000,600000)</td>
</tr>
<tr>
<td>a3</td>
<td>30</td>
<td>(5000,500000,600000)</td>
</tr>
<tr>
<td>a4</td>
<td>70</td>
<td>(2000,400000,500000)</td>
</tr>
<tr>
<td>a5</td>
<td>30</td>
<td>(5000,500000,600000)</td>
</tr>
<tr>
<td>a6</td>
<td>60</td>
<td>(7000,600000,800000)</td>
</tr>
<tr>
<td>a7</td>
<td>30</td>
<td>(5000,500000,600000)</td>
</tr>
</tbody>
</table>

As observe in the figure 5, probability distribution of total cost of project having average of 52699$ as the most possibility cost. This outcome includes expanses value of possibility cost with possibility in each value.

![Figure 5. Preliminarily model of BN for estimating total project cost.](image2)

4-3. Possibilities of modeling relation affiliation cost elements

In section 2, discussed about the importance of determination of positive correlation of cost elements at the time of their combining. In simulation tools of computer (@Risk, crystal ball...) affiliation N component cost (a1,a2,..., ai,...,an) model in form that has shown in figure 6.

![Figure 6. Definition of affiliation in simulation tools.](image3)
Where αi s are coefficient of correlation and 0 ≤ αi ≤ 1 in figure 8. This coefficient determine by expert and for each pair elements of matrix. Definition relation of affiliation cost elements in the simulation method is not compatible with the condition of real world because in practice estimate coefficient of correlation is impractical and unreal for all of the cost elements. In addition, determination coefficient of correlation (αi) for experts will cause the errors in dealing with a number of probability distribution difficult and prone to work. In result, these limitations cause existence many incompatible states with real condition in output of simulation. Moreover with using this simulation, could not model many relations between non-certainties. For example suppose, we want to consider the following affiliation in the estimate total project cost:

<<If the real cost of working package ai is more than real cost of aj, about c $ added to the final project cost>>

This relation will not model with simulation approach. But this work is possible with adding one or more node and appropriate definition of NPT's in Bayessian Networks. For example suppose activities of a2and a3 have the affiliation in high rate and we want, if a2>a3, the total cost increases about 3000$ . We have done this correlation with making a node (Dependencies Management) in kind of Beronlli with two value of 0 and 3000 . Definition of new NPT nodes consist of :<< if a2>a3, number 3000 and if a2≤a3, number 0>>. With exerting this relation between a2and a3, <<project cost at the end of month 2 >> nodes change in form of the sum nodes of a2, a3 and <<Dependencies Management>>. In this management is done the effect of this affiliation in the cost of end of the second month and subsequently total project cost. Average cost will be 53135$ with exerting of this affiliation. As we expected the total project cost has increased with exerting of this correlation. In other word none-exertion of this affiliation caused the low estimate of cost.

4-4. Possibility of analysis of balance between different parameters

Suppose, we have limitation of budget in one or more periods for excreting of special project; and we want to know that in which arrangement reform the cost of periods or previous activities? This work is done with inserting the observation in desirable nodes and helping graphical display of good model that shows the milestones. For instance, suppose for excreting our project till end of third month, doesn’t exist the estimated budget namely 24763$ and only we have 21000$. So we infer the possibilities related to cost of previous activities with inserting this observation, namely:

\[ P(ai| Project cost at the end of month 3=21000), \quad ai=1, 2, 3 \]

In figure 7 has shown the outcome relating to the cost element a3 for previous and after inserting limitation.

![Figure 7](attachment:image.png)

Figure 7. Distribution possibility of cost a3 in forward a) and afterward b) inserting the limitation of budget.

As we see with definition of this scenario change the form of distribution of cost a3 and its parameters. The result of this excretion of scenario announce to manager of project that if according to limitation of mentioned budget, we don't have enough money till end of the third month, it is necessary to finish the activity a3 with 4828 $. In this arrangement excreting the scenario can exist in other parts of costs and expenses. For example, limitation of budget can be for total cost. Excreting this scenario resulted in reform of total cost element. The balance of cost could be more complicate than above state. Sometimes the project management will want to decrease the cost of part of work instead of changing the other variables of projects. For example, we can complete one part in exchange for reducing the limits and quality, increasing time with fewer costs. Moreover the balance of cost-time and cost-quality, analysis of balancing cost-cost is very important in decision making of management. While budget of completing of project is limitable. Sometimes is necessary the cost of other activities become more for completing one activity with fewer cost than estimated value. Model of Bayesian Networks has this analysis. For instance, suppose the budget of total project of our example equal 49000$. In this state the a3 reformed cost equals 5358$. Now,
we want to know, if the cost of activity a6 equals 6700 $ (reducing cost), in that time, the cost of activity a3 how much can increases? With excreting this scenario, the average cost of activity a3 consists of 5452$. Namely, in this condition this activity can increase cost about 94$. Now, this increases permissible cost can be result in decreasing time of this activity or improvement of quality.  

4-5. Diagnostic performance

Here, with using diagnostic characteristic of Bayesian Networks, we present an approach that can be find the factors of increasing the real cost in proportion predicted cost. Note that one of the prominent characteristic of Bayesian Networks and model that here suggested, is characteristic of learning. And in spite of minimum information in estimating preliminary possibilities, in every time observing the secondary possibilities, prepare the possibility inference of preliminary possibilities for the reason of using in next programs. Next step is estimating limits of cost that with occurrence of every factor add to final cost of project (P (CV|Rj)). In the suggested model effect of potential factors examine in form of cost variance (CV). Now in case of we know how much increases of cost (or CV) occur, we can update our knowledge from possibilities of potential factors and in result identify the most possibility of cost variance and weak performance (P(Rj|CV)).

\[ P(\text{Ri}|\text{CV}) = \frac{P(\text{CV|Ri}) \times P(\text{Ri})}{\sum_i P(\text{CV|Ri}) \times P(\text{Ri})} \] (2)

Suppose factors of increase of project example of this paper that applied in modeling of definitive estimate of cost consist of: R1, R2, R3, R4 and R5. For example these risks can be the weak performance of workers, problems related machines, low quality of material and material … So imagine possibility of occurrence of each factor is not available in exact form. So in this state we suppose the possibility of all of them is equal.

\[ P(\text{Ri}) = \frac{1}{5}, \text{ i=1, 2, 3, 4, 5} \] (3)

In other side the deviation of cost resulting from each of above factors estimate as follow:

- R1:CV~U(1000,2000)
- R2:CV~U(400,1500)
- R3:CV~U(0,500)
- R4:CV~U(200,1200)
- R5:CV~U(300,800)

In practice, deviation of cost resulting from identified risks can be every possible distribution (kind of continuous or discrete) or a definite number. Concept of this estimate of cost variance consists of range of increases value of possible cost with possibility of each value on condition of occurrence of related risk. Figure 8 shows methods of modeling above data. Potential factors in nodes << Risks>> and conditional distribution deviation of cost in node <<Cost Variance>> inter the model.

Figure 8: new Bayesian Networks for benefiting of suggested model from diagnostic performance.

Now, we can achieve possibility distribution of cost variance with using above method. This act is called Forward Inference. In this manner that with preliminary possibilities, although is imprecise from chance of occurrence of each risk, possible value of increasing cost and possibility of happening calculate. Outcome of this process is possibility for Backward Inference and achieving to revised probabilities of identified risks. In this arrangement exert the (the last) observation of increasing cost as scenario in possibility distribution of increasing cost and is done the calculation of Bayesian Networks. Figure 9 shows the result of forward conclusion. With running program, display distribution <<cost variance>> with average of 795$. With this concept in which if the possibility of occurrence of total factors of increasing cost be equal, then based on assessment increases cost resulting from each factor, the most possible value for increasing cost of project will be 795$.

Figure 9: Primimilry possibilities and conditional distribution deviation of cost.

In other side suppose cost variance of sample project in reviewed related period (that estimates of possibilities of occurrence and effect of cost variance of factors of cost is done) equal to 1700$. With entering this value as first observation reform the possibilities of factors of increasing cost in form of figure 10. This result can be as a supporter of decision for manager of project in following up reformed action. As observed possibility of occurrence R2 (38%) is more than other factors.
Based on this, it is necessary for benefiting methods for estimating that extract in viewpoint of expert. Judgment information is one of the important sources project consider as a challenge. For that reason, objective information is needed in the process of estimating having risk and is a new activity, completely. So lack of objective information affects the management of projects. Analysis is one of the main concerns of the manager of projects. Due to the nature of the project, the confidence and judgment information of the manager of the project is required. In the study of judgment information and its effect on strategy and research, experts and decision-makers are considered as the source of judgment information. Hence, for developing a model and analyzing it, different databases and information are used, and the model is applied to the current project to verify the results.

4-6. Estimating model

As already mentioned, application of model has shown in simple sample. We have notice that measuring validity this kind of models is different from estimating models that develop in dealing with the routine problems of decision making. Suggested modeling of this paper and similar models develop in dealing with new or difficult situations and by belief and thought of persons in which effect from educating and their experiences. This modeling is tools for thinking, discussion and debate. We apply this modeling for process of new information that can be cause the review and reform with passing time. Section 5-4 certifies these items in this paper. As Check land (2007) and Peed (2009) mentioned, developed models in soft system methodology (SSM) doesn’t consider as a showy from real world, necessarily but insight and expert knowledge of person put in their places and share with others. This kind of model is deliberate creation mind of human and including the relations that must be understand by their creative or user. This perception along with exchanging of thoughts and possibility of discussion and debate consider estimating of this kind of model. This kind of estimating in science said Open – box (White-box) [23, 24]. Based on above said, modeling of this article that forms with soft systems methodology execute in a simple example. This case facilitates repetition and developed method and is a possibility for exchanging mental information of persons and developing model.

5. Conclusion

Cost is one the main parameters of project and affected by uncertainties and estimate and cost analysis is one of the main concerns of the manager of projects. With considering that each project has having risk and is new activity, completely. So lack of objective information in process of estimating a project consider as a challenge. For these cases judgment information is one of the important sources for estimating that extract in viewpoint of expert. Based on this, it is necessary for benefiting methods that can present estimate despite lack of enough information. In last decades, by developing computer technology and being accessible simulation software present methods for cost risk analysis. Although methods based on simulation have theoretical and practical limitation and could not modeling the complicated relations of cause and effect between variables. In this article, in addition a summary reviewing on methodology, using from simulation for become project quantitative cost risks project, present a substitute methods with using Bayesian Networks. Proposed methods in addition become quantitative uncertainties in project cost, having the ability of modeling and measuring different and effective factors on project cost.

Using these developed methods in projects with many complicated can help for better presentation an image (model) of project and in result be effective in process of making decision and project risk management. This article with using a simple example tries to introduce the methods and its abilities. Next studies for developing model can be exerting in following sections:

In relation with application of model and its advantages in proportion to simulation models, can accomplish cases studies and comparable with using information and real scenario.

In relation with developing structure of model, can develop secondary networks for modeling complicated relation (such as unpleasant outcome risk) and then add to model.

In relation with facilitating and improvement efficiency calculating of model can use from “object oriented Bayesian Network” (OOBN)[25].

Corresponding Author:
Hossein Raoofpanah
Raja University Ghazvin, Iran.
Email: h.raoofpanah@yahoo.com

References


3/1/2013