



Stochastic Budget Simulation

Martin Elkjaer

Grundfos A/S, Thorsgade 19C, Iiv., 5000 Odense C, Denmark

Abstract

The purpose of this article is to present a new method for cost estimation. The innovative idea is to combine the conventional calculation method stochastic simulation with basic facets of the successive principle. The purpose of this is to avoid the assessment of dependencies between cost items in the budget. The method is named Stochastic Budget Simulation (SBS), and it is made operational with a software application. The method can be applied to most projects with a simple cost structure at the early stages where uncertainty plays a significant role in estimating the overall cost. The most likely users are planners, project managers or consultants. It is not necessary to understand the calculations, the statistical theory or the simulation technique in order to use the method. However, users should be able to arrange items and overall influences in accordance with the urgent requirement of statistical independence. SBS is a new and radically different way to analyse and evaluate the economic consequences of large-scale projects by quantifying intervals for cost items and using simulation as a tool to represent distributions of the possible costs. © 2000 Elsevier Science Ltd and IPMA. All rights reserved.

Keywords: Cost estimation; Stochastic simulation; Uncertainty analysis

The aim of the method is to establish a reliable and informative economic result based on careful uncertainty analysis and the use of stochastic simulation (synonymous with Monte Carlo simulation). The method combines the conventional Monte Carlo simulation technique with basic facets of the successive principle. [3] The purpose is to avoid statistical correlation between the budget items. This can be done by isolating and separating the overall issues and common dependencies. The successive principle is briefly described in the Appendix.

The purpose of using stochastic simulation is to describe the potential uncertainty of an economic result. The simulation technique makes use of probability distributions to generate a number of the desired overall cost estimate.

SBS may present useful results under certain conditions, which can assist decision-makers in identifying a reliable total cost. However, correct results will also require the use of various other techniques or conditions to ensure that all important matters are included. In addition, the systematic use of evaluation

techniques is required to ensure against evaluation pitfalls. The use is restricted to cost estimates of a very simple structure, while for instance Net Present Value calculations and project durations cannot be dealt with.

1. Context

Many projects are undertaken in a complex environment. Earlier definitions are annulled or at least changed and new situations continually arise. Often there are no reliable data when estimating cost items. At the proposal stage where a feasibility study is usually initiated, the design and demands are still relatively unclear. At this stage it is sensible to consider uncertainties and to use probabilistic range estimation rather than single point estimation, because a probabilistic range reflects the fact that outcomes vary. Stochastic simulation in the form of Monte Carlo simulation is perhaps the most easily usable form of probability analysis.

Current practice, which uses a contingency allowance to cover subsequent design or project changes, is based on deterministic methods or single point estimates. Such methods may serve well under stable conditions, but as the scale and range of variations increases, the utility of this approach is reduced. Many variations require an explicit assessment of uncertainty, and deterministic methods are simply unable to provide this. The situation demands non-deterministic, stochastic methods.

The method can be applied to different types of large-scale projects at the conception stages. Throughout developing projects, software or building projects uncertainty has a crucial impact on the cost components and therefore the total cost. As an example, Stochastic Budget Simulation can be used at the proposal stages of a construction project or in feasibility studies with great effort to evaluate the possible result or total cost.

2. Risk and uncertainty

Before describing the approach of Stochastic Budget Simulation it is necessary to explain the difference between risk and uncertainty. There seems to be some disagreement in the literature regarding the distinction between risk and uncertainty. [4] However, the author finds it suitable to distinguish between the two words, and be careful not to use the words as synonyms. Confusion arises when one regards a subjective risk assessment as an uncertainty analysis.

A risk is a normally unwanted event. It can be identified and quantified through the impact and probability of occurrence. A risk can also be positive, meaning that a risk can be an opportunity to reduce the project cost. Risk can be assessed either objectively or subjectively. Often when no reliable data are available, one has to use subjective judgement to evaluate the consequences of certain risks, which inevitably involves uncertainty. Risks are inevitable in every project and because of risks, uncertainty influences project cost calculations

Risks are therefore integrated into the budget in order to establish a more reliable result. Risks are the overall influences or issues that are common for all the activities or items in the budget. Risks that influence the whole project are named generic risks. These then substitute the traditional contingency allowance in a budget. As the software program cannot handle risks that partly affect some of the items, those risks are neglected. Generic risks are estimated in percentage and multiplied to the sum of the cost items according to normal practice. Generic risks could be price rises, project management, common workforce, common equipment, weather conditions, environmental factors

or team spirit. A risk management procedure can assist in identifying and assessing the potential risks. How this is done lies outside the purpose of this article.

Uncertainty on the other hand is rather more diffuse. In relation to cost estimation, it means that the cost of an item cannot be exactly defined. Uncertainty is an intangible value and is used in case of insufficient knowledge of estimation. Assessment of cost items and generic risks in the budget encompasses uncertainty. Thus the items are regarded as stochastic variables.

Uncertainty analysis should be performed as an integral part of assessing each cost item. Uncertainty analysis is based on the triple estimate using intuitive and subjective judgement. A triple estimate is a way in which to quantify an uncertain value. Uncertainty analysis allows one to obtain quantitative results in the form of confidence intervals. To perform this analysis one must frequently rely on subjective judgement in the absence of information in order to estimate the range of each item in the budget. Using a triple estimate for uncertainty analysis provides planners with an opportunity to quantify the uncertainties involved for the different project items.

3. The approach

This section outlines the approach of SBS. The approach is illustrated below in Fig. 1. It is urgent at this point to emphasise the conditions required for a realistic and reliable economic result. Prior to conducting SBS, the following five steps are recommended.

1. An identification and grouping of all relevant matters with an overall influence upon the project. This requires use of the Work Breakdown Structure (WBS) as well as a consideration of stochastic dependencies. In other words, all cost items need to be identified and included in the budget.
2. A non-biased quantification of conditional cost effects from the above mentioned groups of overall issues. To avoid stochastic dependencies between cost components, a group of generic risks or overall influences is made. The generic risks are assumed to affect all the cost items.
3. The quantification of cost items and generic risks relevant to the inherent uncertainty. A triple estimate is used to quantify the budget items. Careful assessment and systematic judgement are necessary to ensure an accurate total result.
4. The use of algorithms to calculate the total project cost, as well as the local uncertainty for each item. The prime problem here is to avoid stochastic dependencies. If ignored, the results generated will be meaningless.

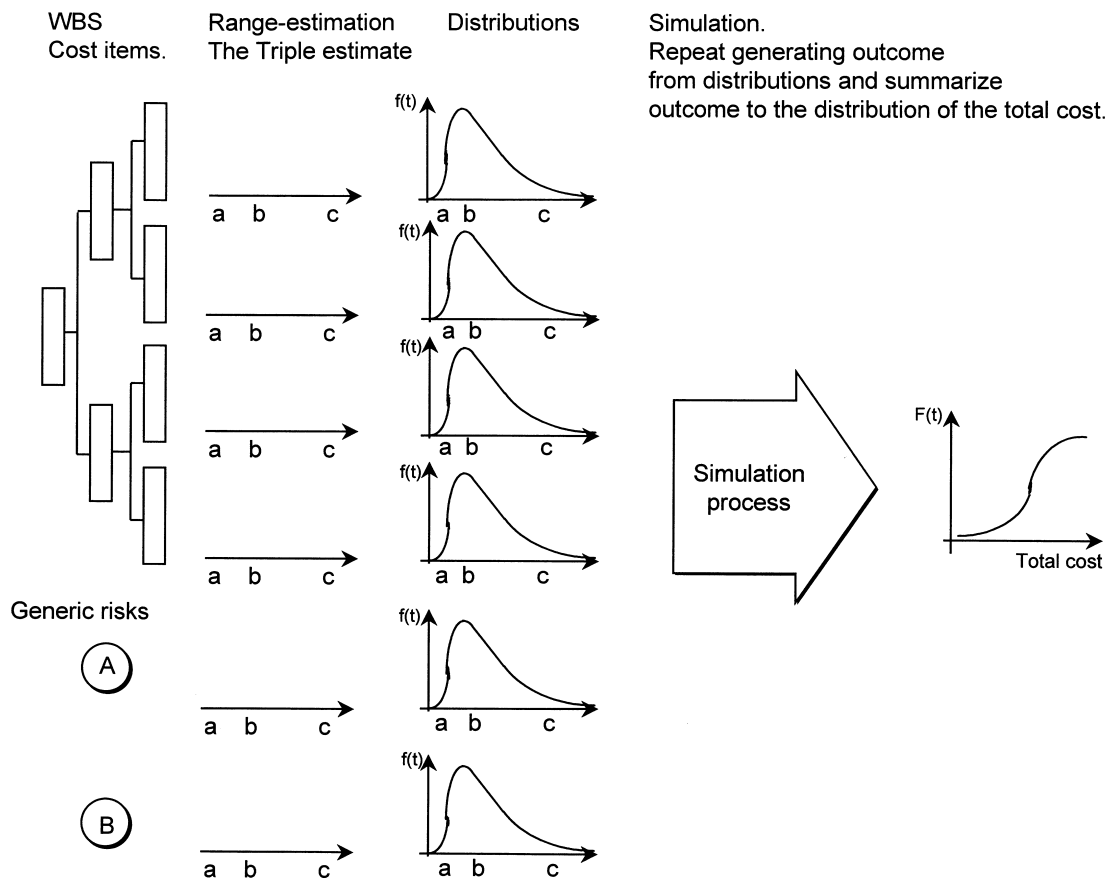


Fig. 1. The approach of Stochastic Budget Simulation.

5. Results must be presented in such a way that project managers can use them to inform stakeholders about the possible economic outcome.

This article focuses on steps 4 and 5 in order to improve the procedures for generating correct and informative results. However, steps 1 to 3 must be carefully handled. Otherwise the mathematical algorithm (the simulation technique) and the idea of grouping common issues seems worthless. Below, the approach for SBS is described and illustrated in Fig. 1.

Initially the project must be structured into a limited number of cost items. These main items are later successively listed according to their priority or effect upon the uncertainty of the total result. The customary specification of costs into hundreds of items allows serious biases to go undetected, such as systematic underestimation. The normal approach generally neglects the importance of focusing on a few vital items and overall influences.

By brainstorming and general experience the planner identifies generic risks and groups these into independent groups. Standard checklists can be valuable to ensure that no matters of major potential effect are omitted. The generic risks must be well-defined in

order to avoid double counting and hidden dependencies in the estimates. The description can include a firm reference definition, which can be used as a common precondition when costs and especially risks are quantified. This works as a baseline for the assessment.

Subsequently each cost item and generic risk is assessed by a triple estimate. At this point generic risks are estimated in percentage. If the estimate for a cost item is cost per unit, for instance £ per m^2 , then the estimate must be multiplied with the value for the unit since the input to the simulation technique has to be monetary values. Generic risks also have to be estimated in cost. As generic risks are regarded as a contingency allowance to the sum of the mean of the cost items, the values are converted into monetary units. As an example, if the sum of the means is 1200, the triple estimate (–10%, 5%, 15%) is transformed to (–120, 60, 180). The range estimation therefore contains three estimates:

A minimum or optimistic value: the lowest possible estimate.

A most likely value: the conventional estimate.

A maximum or pessimistic value: the highest possible estimate.

The actual values for minimum, most likely and maximum can be determined in several ways. The most straightforward method is simply to select the values subjectively, relying upon the expertise of the estimator to determine reasonable values. However, many pitfalls typically violate the result seriously. Sometimes the estimator underestimates the minimum and maximum value. Therefore an approach for careful and systematic assessment is required. This is a significant precondition of a reliable result. A simple and systematic way to estimate the values could be the following:

1. Imagine the lowest possible value.
2. Imagine the highest possible value.
3. Estimate a most likely value between the maximum and minimum value based on experience or reliable information.

After assessing the triple estimate a distribution must be selected. It is possible to choose between an asymmetric triangular function, the Erlang family of distributions or a combination of the possible distributions (see Fig. 2). As described above, the preconditions of structuring the items, identifying the overall influences or generic risks, and systematically quantifying uncertainty are more significant than choosing a correct distribution. However, in order to reduce the difficulty involved in choosing a fair distribution, the software program allows the user to combine all the incorporated distributions.

All cost items are assigned the same distribution due to the functionality of the software program. This reduces the difficulty in choosing a fair distribution for each item. Choosing a correct distribution can be discussed exhaustively, yet it is not the intention here to investigate the choice of a fair distribution.

This topic has been the challenging subject of other papers. [5, 6] The author has included the above men-

tioned distributions, because they are recommended by scientific engineers, [2, 4] and are fairly widespread and familiar.

Simulation can begin once each interval is assigned a probability distribution. The simulation technique consists of the following:

1. A random number between zero and one is generated.
2. By the inverse cumulative distribution a 'random' cost for each item is selected on the basis of the random number between zero and one. It is important to understand that the random number is used to select a value, but the selection process ensures that the frequency with which values are selected conforms to the appropriate distribution.
3. The random cost for each item is summarised to present an overall cost of the project.
4. 1, 2 and 3 are repeated several times to construct a distribution of the total cost.

The simulation process steps through each distribution including the generic risks, determining a single value from the distribution at random. A cost component is then generated within the boundaries of the intervals. The cost components are then added in a conventional way to calculate a total cost for that particular iteration. At the end of each iteration the total cost estimate is recorded prior to repeating the entire process over multiple iterations. Typically 500 iterations are more than enough to produce a result for the total cost. A larger number of iterations gives only a marginal increase in accuracy, and it is of relatively little importance compared with the assessment of the triple estimate. However, the larger the number of iterations, the smoother the graph. This increases visibility during the presentation of the results.

Finally the frequency and cumulative distribution are calculated. These are produced on the basis of a number of iterations for the overall cost. See Fig. 3.

The mean value (μ) and standard deviation (σ) are calculated on the basis of the frequency distribution of

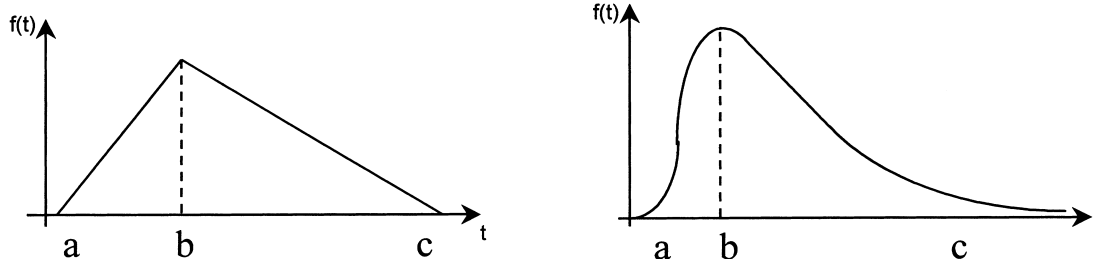


Fig. 2. The figure shows a triangular distribution and an Erlang-5 distribution, which are among the possible underlying distributions for the cost items in the budget. The values a, b and c represent respectively the minimum, most likely and maximum estimate which define the interval.

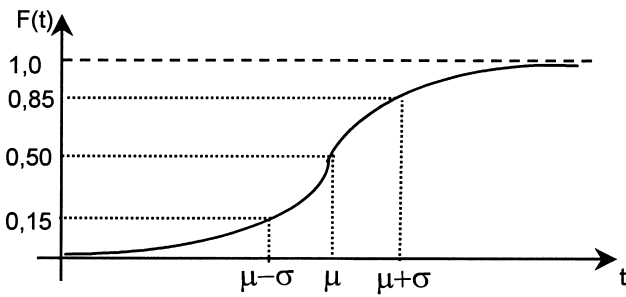


Fig. 3. The figure shows the cumulative distribution for the overall cost, where $F(t)$ represents the probability between $[0,1]$. $\mu \pm \sigma$ indicate the 70% confidence interval. For instance, the cumulative frequency curve can be used to indicate the probability that the total cost will not exceed a particular value. The probability distribution will approximately be a normal distribution according to the central limit theorem.

the total cost. The formulas for mean and standard deviation are listed below:

$$\mu = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$\sigma^2 = \frac{1}{n} \left(\sum_{i=1}^n Y_i^2 - n(\mu)^2 \right)$$

The mean μ is calculated by adding all the values for the total cost (Y_i) and dividing the sum by the number of iterations for the total cost (n). The standard deviation σ is a measure of the spread of the distribution. Due to the application of the simulation technique, the results differ from using the calculation methods in the successive principle. In short, the simulation technique produces a mathematically correct result for the total cost whereas the successive principle produces approximate results.

The SBS is quite different from earlier suggestions for cost simulation. [1, 2, 4] The idea is to combine features from the successive principle with the calculation method stochastic simulation. Conclusively, statistical dependency between cost elements is now treated by the systematic separation of overall influences or generic risks, and the correlation effects are included in an appropriate manner. Range methods and most other similar methods generally neglect important stochastic dependencies, and thus violate statistical laws. The correlation effects are seriously treated as an important contribution to the final result. Common issues and generic risks are therefore identified and estimated as well as the regular items in the budget. The simulation technique ensures that the inherent uncertainty in all items is treated explicitly and in a mathematically cor-

rect manner and transferred to the final distribution through a large number of iterations.

4. Features of the software application

The method Stochastic Budget Simulation is made operational by a software program application based on Excel spreadsheets and Visual Basic. The main feature of the software program is to handle the stochastic simulation. The software program makes it possible to perform a sensitivity analysis, as it is possible to change the parameters (the minimum estimate, the most likely estimate and the maximum estimate) for specific cost items. This might be done if the calculator assesses that for instance the pessimistic parameter is too low. With a sensibility analysis, it is possible to analyse the outcome of the simulation or the consequences for the overall cost by changing the value of cost components.

The software program also allows the user to identify the cost items, which carry most uncertainty. It is optional for the user to specify cost items in order to receive a more reliable result. The simulation process can be performed any time the user finds it appropriate. The user does not have to understand the mathematical theory to use the software program.

5. An example

Although the primary objective of this paper is to present a new approach to calculating the overall cost of any project in the conception phases, the following example will be used to illustrate the operational use and features of SBS. The example is based on a fictive developing software project. The budget is therefore not complete and the estimates do not reflect realistic values. Due to a better comprehension of the application of SBS the spreadsheets are visualised.

After having identified the cost items and generic risks, a triple estimate for each item is calculated. The items and their estimates are entered in the main sheet below (see Fig. 4). When the user types the estimates for the items, it is possible to protect the main items. This feature secures that headings for groups are not estimated, if they are wanted in the budget.

The sum of the means for the cost items is \$1451000. The generic risks are firstly estimated in percentage, and subsequently added to \$1451000. For instance the triple estimate for project management is (−15%, 0%, 30%) which equals (−217,6, 0, 435,3).

The mean and standard deviation are calculated for each item by using the approximate formulas from the successive principle. The purpose of these calculations is to rank the ten items that contribute most to the

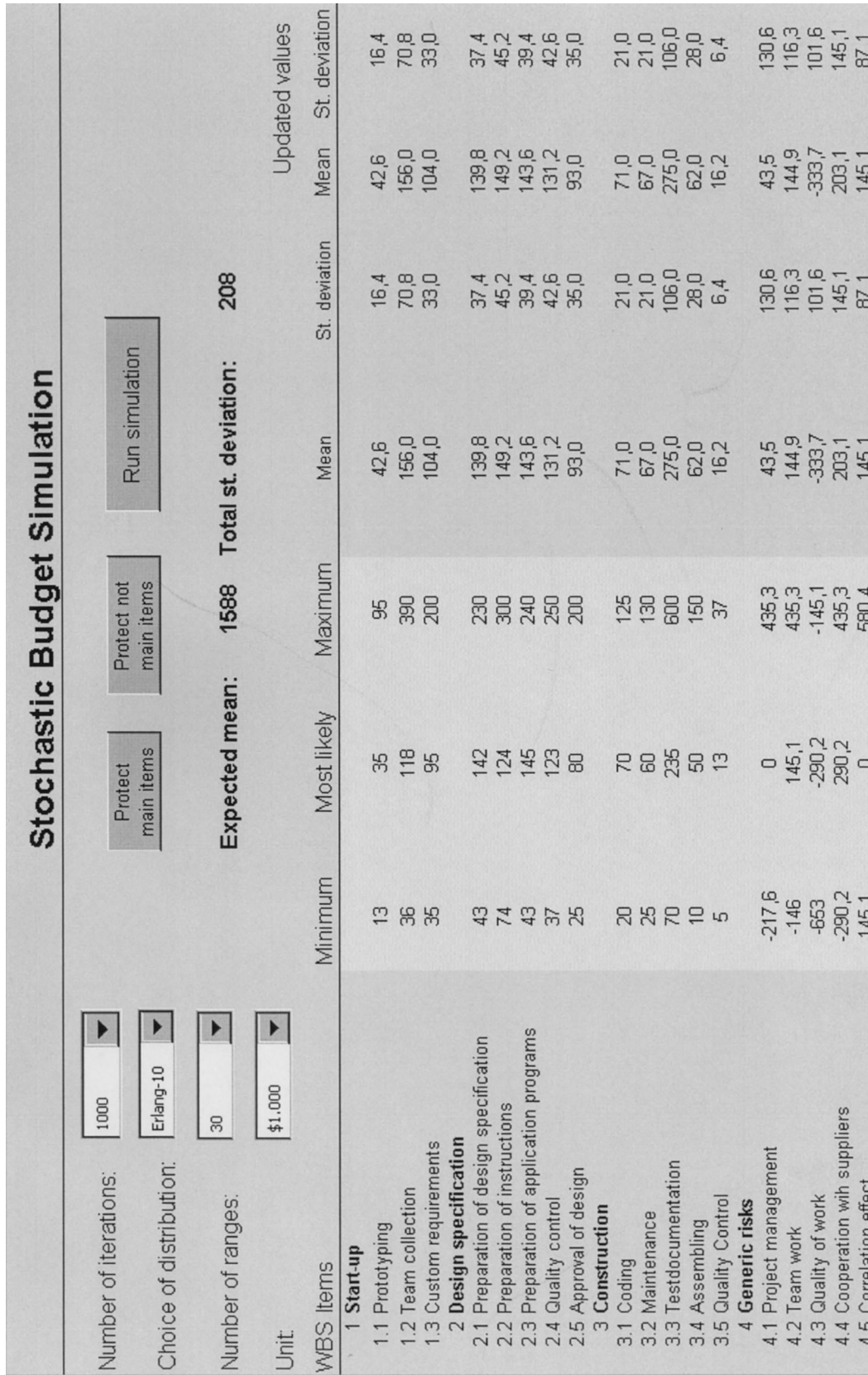


Fig. 4. Main sheet.

Update list		Priority list	Specification
No.	WBS	Items	Relative deviation
1	4.4	Cooperation with suppliers	21,7%
2	4.1	Project management	17,6%
3	4.2	Team work	14,0%
4	3.3	Testdocumentation	11,6%
5	4.3	Quality of work	10,7%
6	4.5	Correlation effect	7,8%
7	1.2	Team collection	5,2%
8	2.2	Preparation of instructions	2,1%
9	2.4	Quality control	1,9%
10	2.3	Preparation of application programs	1,6%

Fig. 5. Prio.

total uncertainty. These are therefore automatically placed in the priority list, which can be updated at anytime (see Fig. 5).

The priority list calculates a comparable effect of each uncertain cost item or generic risk upon the uncertainty of the total result measured by the standard deviation. The list indicates how important the local item is compared to others. The user can then specify an item into sub-items. The value for the relative deviation in the priority list is the indicator for further specification into independent groups. The relative deviation is calculated by dividing the local standard deviation for each item with the total standard deviation. In the example co-operation with suppliers is the most uncertain factor, and in order to reduce the total uncertainty of the project, it must be further

specified. This can be done by marking the item with the cursor, and then pushing the ‘Specification’ button. A new sheet is then ready for detailed analysis into sub-items.

A specification usually results in different values for the mean and deviation, hence the values are updated and replaced in the main sheet (see Fig. 4). After each specification the priority list is normally updated.

The main sheet for the triple estimates also contains other facilities. The user has to determine the number of iterations or how many times the total cost must be calculated. In this example 1000 iterations are chosen, which is more than sufficient for an acceptable result. The monetary unit for the items must also be determined, and here \$1000 is selected.

Due to the visualisation of the frequency graph a number of ranges must also be specified. This makes it possible to count the number of iterations in specific intervals. The number of ranges affects the visualisation of the frequency distribution. The more ranges are chosen the smoother the illustration of the graph.

Finally, the distribution can be selected. The user is free to choose any of the possible distributions in which the user has most confidence. Even though the choice of distribution type has an influence on the final results, the preconditions outlined in steps 1 to 3 are more significant. In Fig. 4 an Erlang–10 distribution for each item is preferred.

Probability and cumulative distribution

Project: SAP2002 Manager: S. Michael Date: 2000-02-16 Client: Int. Data

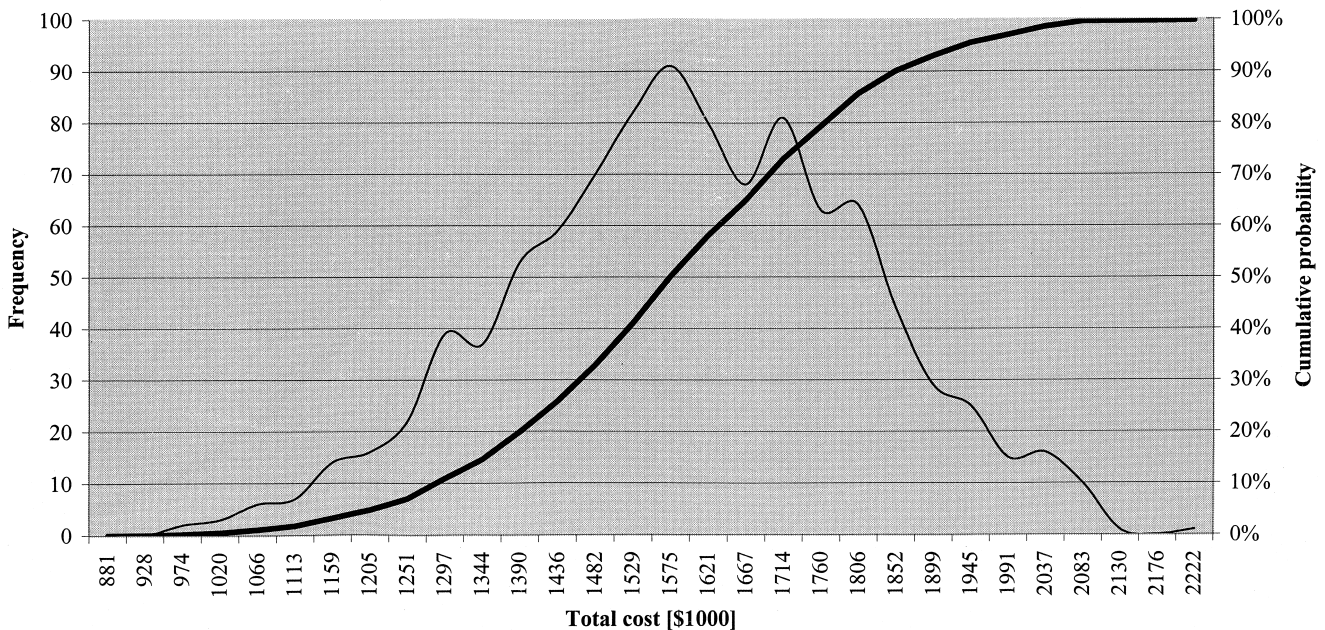


Fig. 6. Graph.

Table 1

	Triangular	Erlang-10	Combination
Expected mean	1731	1588	1639
Expected standard deviation	234	208	229

Then the budget is ready for stochastic simulation. By activating the button 'Run simulation', the total expected mean and standard deviation are calculated respectively to \$1 588 000 and \$208 000. A probability and cumulative distribution is also generated as illustrated in Fig. 6. Using the cumulative distribution, decision-makers can make decisions based on reliable mathematical documentation for the final cost.

The graph illustrates the calculated mean and standard deviation for the total cost. It is particularly important to notice that these are calculated on the basis of the outcome of the simulation. The cumulative distribution can be used to indicate the chances that the total costs do not exceed a particular value. As an example, Fig. 6 shows that there is a 70% probability that the total cost will be less than about \$1 750 000. If the investor has a specific amount of money, he can evaluate the success for implementing the project within the budget limits. For example, a specified investment of \$1 500 000 has a 30% chance of staying within the budget cost (see Fig. 6). These conclusions are dependent upon a sufficient analysis and successful completion of the mentioned five steps.

Although the selection of a correct distribution is not very significant compared to the preconditions, the results for the total cost will differ depending on the selected distributions. Table 1 illustrates the values for the total cost for selected distributions on the basis of the fictive example. The values are in \$1000.

There is a difference of approximate 8% and 11% respectively between the highest and lowest value for the expected mean and standard deviation. Even though research indicates that the Erlang family of distributions expresses relatively reliable uncertainty estimations, the author recommends choosing a combination of all the included distributions.

6. The results

By using Stochastic Budget Simulation planners, decision-makers are able to make decisions based on a mathematically exact distribution instead of approximate algorithms. If the preconditions are well performed the total distribution might show the actual costs. The distribution of the total costs presents the expected mean and standard deviation and subsequently establishes a confidence interval. The distri-

bution further indicates the probability that costs will not exceed a particular value. The priority list enables project managers to focus on the most important items that need further specification in order to reduce the overall uncertainty.

After quantification by use of the triple estimate, the distribution of the total cost is dependent on the type of underlying distributions and the amount of iterations. As seen above, a triangular distribution and an Erlang distribution give different results for the approximate normal distribution of the total cost. It cannot be concluded which distribution is the most appropriate, because the final cost of a project is naturally not known. As seen, it is relatively important for analysing the expected total cost which underlying distribution is used, if the preconditions are well executed. However, it should be noted that the k value for the Erlang- k distribution for the mean value of the total cost is not of importance, but the standard deviation decreases as the k value increases in accordance with the theory.

Instead of using direct approximate algorithms to calculate the overall cost, this method performs an exact calculation using the stochastic simulation technique.

7. Conclusions

Most projects are conducted in a changing environment, which makes the analysis of the project economy in the early stages quite difficult. It is necessary to study the uncertainties involved in the project and to let the economic result reflect the possible total costs. By using a probabilistic approach by including distributions for each item in the budget, decision-makers will have an analytical tool with which to evaluate the most likely total cost. This is done with the use of stochastic simulation technique. By using facets of the successive principle, the users do not have to worry about correlation between the cost items as common dependencies are isolated and separately estimated.

SBS is an operational tool for planners, which is easy to use and quickly presents an overview of the total cost. Furthermore, the estimator can conduct a sensibility study and focus on the items with most local uncertainty compared to the overall uncertainty. A specification of the items can be performed to ensure a more accurate result. SBS may improve project results subject to the condition that cost items and generic risks are properly identified and evaluated.

The author does not claim that the method is the ultimate tool to present a reliable economic result at the early project stages, but the author has introduced an alternative method, which is a good example of a future application.

Appendix A. The successive principle and its scope

The successive principle is a tool for project managers and decisions-makers who require the inclusion not only of regular cost items, but also of all the relevant fuzzy factors affecting their work. The principle is used in most private firms and public companies to support and facilitate estimations, allowance and guarantee decisions, scheduling, commercial risk analysis as well during start-up and teambuilding phases of new ventures. The applications and benefits are primarily the following three: Firstly, it is possible to make very realistic budget estimates, project durations etc., and thus largely eliminate overruns and other unpleasant surprises. This can even be done at a very early phase of the plans. Secondly, as part of a built-in ranking process, the responsible managers are given a prioritised list of critical items or activities that contribute strongly to uncertainty in the project. Thirdly, the mutual understanding of the aims and characteristics of a given project or program are radically improved among the involved key persons, thus also improving the important teambuilding process in the project group. The method basically involves listing all factors of importance, not only the physical and formal items, but also the fuzzy and sensible matters, and openly and correctly to control and handle uncertainty and even to consider uncertainty as an existing aspect in planning and managing. For reasons of overview and rapid performance the successive principle uses a top down approach starting with the main items and successively developing a work break down structure for those items where uncertainty is highly critical. Due to the complexity of projects, it is considered essential to perform the analysis jointly via a group of key persons. This also has positive side effects such as increased consensus and strengthened team building. The general procedure outlined:

1. A group of key persons gather. The first task of the group is to thoroughly discuss the tasks, pre-conditions and objectives.
2. All general sources of potential uncertainty are identified, organised in groups and defined according to relevant sub routines.
3. A set of main items or activities is chosen, and a triple estimate for each item is made. One or more generic risks or overall influences are added, based upon potential deviations from the reference defined in step 2.
4. Direct approximate procedures are performed using statistical rules. The mean and standard de-

viation of the total is calculated, and the priority list is created. The formulas for the local mean and local standard deviation are respectively $(\text{min.} + 3 * \text{most likely} + \text{max.}) / 5$ and $(\text{max.} - \text{min.}) / 5$. (By comparison in using stochastic simulation, these values are added respectively to the total mean and standard deviation for the overall cost.)

5. The most critical items are successively detailed. The guidance in this detailing process is the priority list, which indicates the relative importance of the individual item to the total uncertainty. This continues until a reasonable minimum of uncertainty is reached.
6. The results of this procedure are a highly mean value and a 'top ten list' with the remaining major items or risks that consist of most uncertainty. This list is typically followed up by an action plan suggested by the analysis group.

References

- [1] Cooper DF, MacDonald DH, Chapman CB. Risk analysis of a construction cost estimate. *Journal of Project Management* 1985;3(3):141–9.
- [2] Flanagan R, Kendall A, Norman G, Robinson GD. Life cycle costing and risk management. *Construction Management and Economics* 1987;5:53–71.
- [3] Lichtenberg S. New project management principles for the conception stage. *Proceedings, INTERNET 88 and Journal of Project Management* 1989;7(1):46–51.
- [4] Newton S. Methods of analysing risk exposure in the cost estimates of high quality offices. *Construction Management and Economics* 1992;10:431–49.
- [5] Wall DM. Distributions and correlations in Monte Carlo simulation. *Journal of Construction Management and Economics* 1997;15:241–58.
- [6] Touran A, Wiser EP. Monte Carlo technique with correlated random variables. *Journal of Construction Engineering and Management* 1992;188(2):258–72.
- [7] Elkjaer M. *Project management of cost and risk* (English title). Institute of Planning, DTU, DK, 1998.



Martin Elkjaer graduated in April 1998 from the Technical University of Denmark. He works as a management consultant for Pricewaterhouse Coopers in Copenhagen, Denmark. This article is based on Mr. Elkjaer's report "Project management of cost and risk" [7] which concluded his Master of Science in Engineering (Planning and Technology Management). Requests or questions can be forwarded to martin_elkjaer@hotmail.com