DETERMINING THE CRITICAL ACTIVITIES IN A COMPLEX PROJECT USING SIMULATIONS

Deyan Mihaylov

University of Economics-Varna
e-mail: dgmihaylov@ue-varna.bg
Bulgaria

Abstract: The subject of this article is to represent a simulation model which can determine the activities, which can directly impact the duration of complex projects. It is assumed that the duration of every activity is a random variable. The model determines all the paths in the project diagram and the relative frequencies of occurrences of critical paths and activities. The critical activities could affect the duration of the whole project and require more attention. Variants in which the distributions are PERT-beta, triangular and uniform are played out. The experiment shows that the type of distribution has great impact on the relative frequencies.

Key words: Critical Activity, Random variables, Simulations.

1. INTRODUCTION

The Program Evaluation and Review Technique (PERT) was developed during the 1950s. The key concept is that a complex task or project can be represented as a set of activities which are logically interdependent. The structure of the whole project is a network which comprises of events and activities. The activities can run sequentially or in parallel. The execution of activities begins after the project's starting point. When all activities are completed, the finish point occurs. Every chain of activities between the start point and finish point is called a path. Every activity has a duration time. The path duration is a sum of the durations of activities included in this path. The longest path is called a critical path [3].

One of the goals of analysing the network plan is to determine the activities which have the greatest influence on the duration of the whole task. These activities are called critical activities. Their identification enables a more effective control of the project [5]. If the duration of the activities is random, the critical path can only be determined stochastically. Theoretically several paths can be critical, hence every activity can directly impact the duration of the project in different cases.

According to Bellman's Principle of Optimality, Vilenkin [9] shows that the duration of every path is the sum of the activity durations included in them and the duration of the critical path is the maximum of the durations of all the paths. Pursuant to the Central Limit Theorem he accepts that a single path's duration has a normal distribution. This statement could be disputed because the activities included in a path are not numerous. For example, Oleinikova and Kravets [7] use Beta distribution to approximate the sum of several random Beta variables.

Obtaining the theoretical distribution law is a very difficult task. The distribution of a sum of random variables can be derived in two ways. The first one is to solve a convolution integral. The second one is to use the characteristic functions. Both methods are very complex. The same activity can be part of several paths, so the durations of different paths are not independent. Hence, the law of distribution of their maximum is their joint multidimensional nonseparable function. So it is reasonable to use simulation techniques to investigate the network model.

The purpose of this article is to explain a way to solve the problem of determining the direct impact of every activity on the project's duration. The constraints of this study are:

1. The activity has a direct impact on the project duration when every small increase of activity duration causes an increase of project duration. This means that the total slack of this activity is equal to zero.

2. It is assumed that the duration of every activity is a random variable and that all random variables are mutually independent.

Historically the first model of activity duration in PERT uses a special case of beta distribution called PERT-beta distribution. However this is based on the intuition of the method's authors and not on rigorous proof [1]. Recently different distributions have been used for modeling activity time. Hajdu & Bokor [2] show that the use of uniform and triangular distribution has a small effect on the project duration. It would be good to compare how different distributions impact critical activities. In this study PERT-beta, triangular and uniform distributions are used in the simulation model.

2. SIMULATION MODEL OF A NETWORK PLAN

Let the complex project consist of n activities A_1, A_2, …, A_j, …, A_n. The formal representation of every activity is

$$A_j = \{a_j, b_j, m_j, Pr_j\} \tag{1}$$

where a_j is the optimistic time;

b_j is the pessimistic time;

m_j is the most likely time;

Pr_j is the set of predecessors of A_j.

Let ξ_j denote the duration of activity A_j and according to classical view ξ_j is a random variable [4]. The variants of models are:

1) Every ξ_j has PERT-beta distribution.

The mean, variance and mode of ξ_j are:

$$Mean_j = \frac{a_j + 4m_j + b_j}{6} \tag{2}$$

$$Var_j = \frac{(b_j - a_j)^2}{36} \tag{3}$$

$$Mode_j = m_j \tag{4}$$

The four-parameter Beta distribution $B_j(u_j, v_j, a_j, b_j)$ is defined by four parameters: u_j and v_j are the standard Beta shape parameters; a_j and b_j are the minimum and maximum bounds of the distribution.

The PDF of $B(u_j, v_j, a_j, b_j)$ is

$$f_j(t) = \begin{cases} \dfrac{(t - a_j)^{u_j - 1}(b_j - t)^{v_j - 1}}{\mathrm{B}(u_j, v_j)(b_j - a_j)^{u_j + v_j - 1}} & , t \in [a_j, b_j] \\ 0 & , t \notin [a_j, b_j] \end{cases} \tag{5}$$

In (5) $B(u_j, v_j)$ is the well known beta-function.

The parameters a_j, b_j and m_j make it possible to obtain the shape parameters [6]:

$$u_j = \frac{(Mean_j - a_j)(2m_j - a_j - b_j)}{(m_j - Mean_j)(b_j - a_j)} = \frac{4m_j + b_j - 5a_j}{b_j - a_j} \tag{6}$$

$$v_j = \frac{u_j(b_j - Mean_j)(2m_j - a_j - b_j)}{(Mean_j - a_j)} = \frac{5b_j - a_j - 4m_j}{b_j - a_j} \tag{7}$$

Using (6) and (7) makes it possible to generate random variables which simulate the duration ξ_j of every activity A_j.

2) All ξ_j? are uniform distributed.

Let the random variables be between the optimistic and pessimistic times. Hence, the uniform distribution is $\xi_j \sim U(a_j, b_j)$ and some of its characteristics and PDF are

$$Mean_j = \frac{a_j + b_j}{2} \tag{8}$$

$$Var_j = \frac{(b_j - a_j)^2}{12} \tag{9}$$

$$f_j(t) = \begin{cases} \dfrac{1}{b_j - a_j} & , t \in [a_j, b_j] \\ 0 & , t \notin [a_j, b_j] \end{cases} \tag{10}$$

The mode does not exist.

3) All ξ_j are triangular distributed.

The random variables are between the optimistic and pessimistic times and the most likely value is equal to m_j. Hence, the triangular distribution is $\xi_j \sim T(a_j, b_j, m_j)$ and its characteristics and PDF are

$$Mean_j = \frac{a_j + b_j + m_j}{3} \tag{11}$$

$$Var_j = \frac{a_j^2 + b_j^2 + m_j^2 - a_j b_j - a_j m_j - b_j m_j}{18} \tag{12}$$

$$Mode_j = m_j \tag{13}$$

$$f_j(t) = \begin{cases} 0 & t < a_j \\ \dfrac{2(t - a_j)}{(b_j - a_j)(c_j - a_j)} & a_j \le t < m_j \\ \dfrac{2(b_j - t)}{(b_j - a_j)(b_j - c_j)} & m_j \le t < b_j \\ 0 & b_j \le t \end{cases} \tag{14}$$

Let the critical path be determined. The difference between the duration of the critical path and the duration of one of the non-critical paths is called the path float of the non-critical path in question. The path float denotes the amount of time by which the actual completion time of an activity on the path in question can exceed its earliest completion time without affecting the earliest start or occurrence time of any activity or event on the network critical path [5]. Hence, the activities included in the critical path are critical activities and activities which do not belong to the critical path are non-critical.

The modeling of the plan includes the following steps:

1. Determining all the paths $Path_i$, $i = 1,2,3,...,m$.
2. Generating the durations of every activity.
3. Calculating the duration of every path.
4. Determining the critical path (which has the longest duration).
5. Determining the critical activities, i.e. activities included in the critical path.

The sequence described above is repeated N times. Let M_j be the number of simulations in which A_j is a critical activity. The relative frequency

$$P_{Cr}\left(A_j\right)=\frac{M_j}{N} \tag{15}$$

measures the impact that every delaying of A_j has on the duration of the project.

It should be noted that the activity duration usually is an integer (hours, days, weeks). Hence, it is useful to investigate the case when the obtained random values are rounded.

3. EXAMPLE

According to the aforementioned, some statistical experiments were done. The example is based on a fictive project. The parameters of activities are shown in Table 1.

Table 1. Activity characteristics of a fictive project

Activity	**Time Estimated**			**Predecessors**
	Optimistic time (a **)**	**Pessimistic Time (** b **)**	**Most Likely Time (** m **)**	
1	**2**	**3**	**4**	**5**
A01	3	7	6	None
A02	6	11	7	None
A03	3	9	8	A01
A04	7	10	9	A01
A05	2	5	3	A02
A06	7	10	8	A02
A07	4	9	8	A03
A08	5	10	8	A04
A09	1	3	2	A05
A10	2	6	4	A05
A11	6	9	7	A06
A12	5	10	9	A07, A08, A09
A13	7	10	9	A10, A11

A diagram of the project in “Activity on Arc” form is shown in Figure 1.

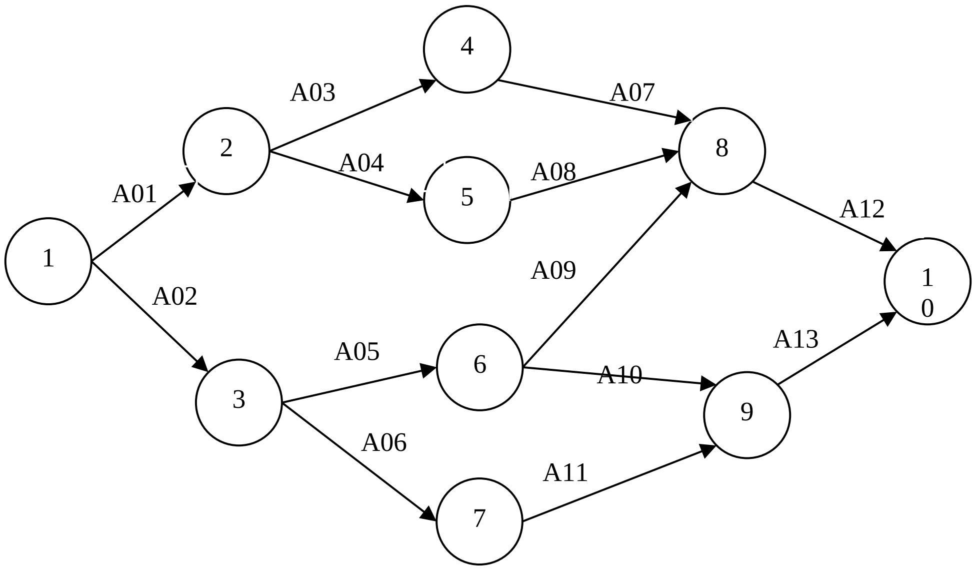


Fig. 1. "Activity on Arc" Diagram.

There are five paths in the model. The sequences of activities in the paths are

Path1 A01—A03—A07—A12;
Path2 A01—A04—A08—A12;
Path3 A02—A05—A09—A12;
Path4 A02—A05—A10—A13;
Path5 A02—A06—A11—A13.

The PDFs of random variables with PERT-beta, triangular and uniform distributions associated with activity A05 are shown in Figure 2.

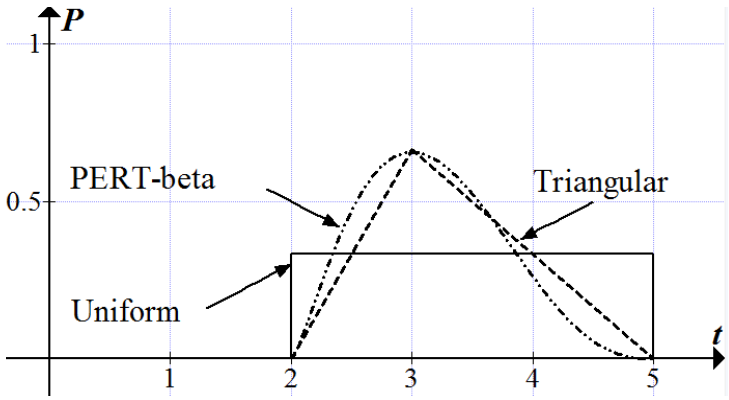


Fig. 2. Pert-beta, triangular and uniform distributions.

Every distribution was examined in two variants. The random values which were used in the first one were obtained from a random generator without rounding (NR-signed). In the second one the random values were rounded (R-signed). 100 000 simulations were done for each experiment. The relative frequencies with which the corresponding path is critical are shown in Table 2.

Table 2. Critical Paths Frequencies

Paths	**Pert-Beta**		**Uniform**		**Triangular**	
	NR	**R**	**NR**	**R**	**NR**	**R**
1	**2**	**3**	**4**	**5**	**6**	**7**
Path1	0.036	0.084	0.005	0.011	0.006	0.017
Path2	0.329	0.428	0.083	0.121	0.130	0.192
Path3	0	0	0	0	0	0
Path4	0	0	0	0	0	0
Path5	0.635	0.707	0.912	0.925	0.864	0.897

The relative frequencies for the activities are shown in Table 3.

Table 3. Relative Frequencies of Activities

Activity	**Pert-Beta**		**Uniform**		**Triangular**	
	NR	**R**	**NR**	**R**	**NR**	**R**
1	**2**	**3**	**4**	**5**	**6**	**7**
A01	0.365	0.469	0.088	0.129	0.136	0.201
A02	0.635	0.707	0.912	0.925	0.864	0.897
A03	0.036	0.084	0.005	0.011	0.006	0.017
A04	0.329	0.428	0.083	0.121	0.130	0.192
A05	0	0	0	0	0	0
A06	0.635	0.707	0.912	0.925	0.864	0.897
A07	0.036	0.084	0.005	0.011	0.006	0.017
A08	0.329	0.428	0.083	0.121	0.130	0.192
A09	0	0	0	0	0	0
A10	0	0	0	0	0	0
A11	0.635	0.707	0.912	0.925	0.864	0.897
A12	0.365	0.469	0.088	0.129	0.136	0.201
A13	0.635	0.707	0.912	0.925	0.864	0.897

It is possible for some paths to be critical at the same time, so the events "Path №… is critical" can be compatible. Hence, the sum of relevant frequencies (probabilities) generally is not equal to 1. But the total sums in Table 2, columns 2, 4 and 6 are equal to 1. This is an indication that in each of these simulations only one path is critical. On the other hand the sums of data in Table 2, columns 3, 5 and 7 are more than 1 and this shows that during some simulations several paths were critical at the same time. The explanation is very simple. The frequencies in columns 2, 4 and 6 are caused by the processing of absolutely continuous random variables. The probability that an absolutely continuous random variable takes a specific value tends to zero. Hence, it is almost impossible for two or more path durations modeled by an absolutely continuous random variable to take the same value, moreover this value to be a maximum, i.e. critical.

But projects can have multiple critical paths. This possibility can be modeled by rounding the activity time. The rounded random variable is actually a discrete variable. It takes a finite number of values and the possibility that another random variable takes the same value exists, therefore the probability that several paths can be critical at the same time can be positive. It seems that the second approach (modeling by rounded random variables) is more adequate to reality. The variables are rounded to integer values in the presented example, but the step of rounding can be different.

Relative frequencies of activities (Table 3) represent the grade of impact of the corresponding activity on the duration of the project. The grade of impact is proportional to the probability that the activity lies on a critical path. Let an activity lie on several paths which can be critical separately or at the same time. The assumption that several paths can be critical at the same time leads to the conclusion that the probability of the activity in question generally is not a sum of the probabilities of all the paths on which this activity lies. For example, activity A01 lies on Path 1 and Path 2. Both paths can be critical. So, the possibility that both paths are critical at the same time exists. The sum of their probabilities when the durations are modeled by rounded variables is more than the probability of A01, but when the variables are not rounded, this sum is strictly equal to the probability of A01.

It can be seen that relative frequencies generally depend on the type of distributions. It seems paradoxical, but the difference between PERT-beta and triangular distributions, which look very close, is greater than the difference between triangular and uniform distributions, which do not look similar. The explanation can be found in the difference between the characteristics of the random variables. The means, variances and standard deviations of activity time distributions and paths are shown in Tables 4 and 5.

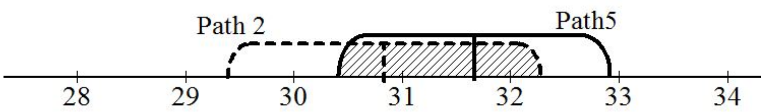
Table 4. Characteristics of Activities

	a	*b*	*m*	**PERT-beta**			**Triangular**			**Uniform**		
				Mean	**Var**	**Sd**	**Mean**	**Var**	**Sd**	**Mean**	**Var**	**Sd**
A01	3	7	6	5.67	0.44	0.67	5.33	0.72	0.85	5	1.33	1.15
A02	6	11	7	7.5	0.69	0.83	8	1.17	1.08	8.5	2.08	1.44
A03	3	9	8	7.33	1	1	6.67	1.72	1.31	6	3	1.73
A04	7	10	9	8.83	0.25	0.5	8.67	0.39	0.62	8.5	0.75	0.87
A05	2	5	3	3.17	0.25	0.5	3.33	0.39	0.62	3.5	0.75	0.87
A06	7	10	8	8.17	0.25	0.5	8.33	0.39	0.62	8.5	0.75	0.87
A07	4	9	8	7.5	0.69	0.83	7	1.17	1.08	6.5	2.08	1.44
A08	5	10	8	7.83	0.69	0.83	7.67	1.06	1.03	7.5	2.08	1.44
A09	1	3	2	2	0.11	0.33	2	0.17	0.41	2	0.33	0.58
A10	2	6	4	4	0.44	0.67	4	0.67	0.82	4	1.33	1.15
A11	6	9	7	7.17	0.25	0.5	7.33	0.39	0.62	7.5	0.75	0.87
A12	5	10	9	8.5	0.69	0.83	8	1.17	1.08	7.5	2.08	1.44
A13	7	10	9	8.83	0.25	0.5	8.67	0.39	0.62	8.5	0.75	0.87

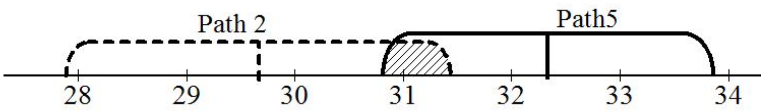
Table 5. Characteristics of Paths

	PERT-beta			**Triangular**			**Uniform**		
	Mean	**Var**	**Sd**	**Mean**	**Var**	**Sd**	**Mean**	**Var**	**Sd**
Path1	29	2.83	1.68	27	4.78	2.19	25	8.5	2.92
Path2	30.83	2.08	1.44	29.67	3.33	1.83	29	6.25	2.5
Path3	21.17	1.75	1.32	21.33	2.89	1.7	22	5.25	2.29
Path4	23.5	1.64	1.28	24	2.61	1.62	25	4.92	2.22
Path5	31.67	1.44	1.2	32.33	2.33	1.53	33	4.33	2.08

Let the cases of non-rounded PERT-beta and triangular distributions be analyzed. In both cases Path 5 has a greater mean of duration than Path 2. But the means and the standard deviations are different. This causes a different probability that the duration of Path 2 is greater than the duration of Path 5 i.e. that Path 2 is critical. The shaded areas in Figure 3 show the overlapping of intervals (Mean-Sd, Mean+Sd) in both cases. It can be seen that in the case of triangular distributions the probability that Path 2 is critical is less than in the case of PERT-beta distributions. This corresponds to the results in Table 2.



a) PERT-beta distributions.



b) Triangular distributions

Fig. 3. The intervals (*Mean-Sd, Mean+Sd*) *for Path 2 and Path 5 durations.*

The hypothesis that the impact on probabilities depends on characteristics of random variables but not on the shape of distributions can be examined. Let the uniform distributions which correspond to the following activity durations be used:

The lower and upper bounds are

$$Low = \frac{a_j + 4m_j + b_j - \sqrt{3}\left(b_j - a_j\right)}{6};$$

$$Upp = \frac{a_j + 4m_j + b_j + \sqrt{3}\left(b_j - a_j\right)}{6}.$$

The durations of activities are random variables

$$\xi_j \sim \left(Low, Upp\right) \tag{16}$$

and their characteristics can be obtained using formulas (8) and (9):

$$Mean_j = \frac{Low + Upp}{2} = \frac{a_j + 4m_j + b_j}{6} \tag{17}$$

$$Var_j = \frac{\left(Upp - Low\right)^2}{12} = \frac{\left(b_j - a_j\right)^2}{36} \tag{18}$$

It should be emphasized that the bounds of distributions (16) do not coincide at all with the pessimistic and optimistic times of relevant activities. But characteristics (17) and (18) coincide to the characteristics of the corresponding PERT-beta distributions which are obtained using formulas (2) and (3).

The model has been run using distributions (16). A comparison of the result of the simulations with a PERT-beta model is presented in Table 6. It should be noted that the shapes of the PERT-beta distribution and the uniform distribution are very different.

Table 6. Comparison of a PERT-beta model and a modified uniform model.

Paths	PERT-beta		Modified Uniform	
	NR	**R**	**NR**	**R**
1	**4**	**5**	**4**	**5**
Path1	0.036	0.084	0.039	0.089
Path2	0.329	0.428	0.311	0.414
Path3	0	0	0	0
Path4	0	0	0	0
Path5	0.635	0.707	0.650	0.717

It can be seen that the differences between the models are very small. It is more important that the means and variances are similar rather than the shapes of distributions.

4. CONCLUSION

This article presents a simulation model which allows determining the probability that a path or an activity can be critical in a complex project. The number of simulations is big, so it can be claimed that the obtained relative frequencies are approximately equal to the relevant probabilities.

The example shows that the type of the distributions generally **has great effect** on determining whether activities are critical or not. It turns out that this choice can have a small effect on some parameters (for example on the project duration) but a great impact on others. Unfortunately, the input data concerning activities usually has a limited number of estimators. This makes choosing a distribution a difficult task.

The simulation model was made using a platform with the following specs: Intel® i3 Core™ i3-3130M CPU@2,6GHz; RAM 4 GB; Windows 10 Education, 64-bit; R-language Version R-x64 3.4.2 [8]. The running time of 100 000 simulations take about 4-5 min.

REFERENCES

[1] Clark, C. E. Letter to the Editor—The PERT Model for the Distribution of an Activity Time. *Operations Research* , **3** (vol. 10), 1962, pp. 405-406.

[2] Hajdu, M., & Bokor, O. The Effects of Different Activity Distributions on Project Duration in PERT Networks. *Procedia - Social and Behavioral Sciences* , 119, 2014, pp. 766-775.

[3] Kerzner, H. *Project Management : a Systems Approach to Planning, Scheduling, and Controlling.* John Wiley & Sons, Hoboken, New Jersey , USA, 2017.

[4] Malcolm, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W. Application of a Technique for Research and Development Program Evaluation. *Operations Research* , **5** (vol.7), 1959, pp. 646–669.

[5] Moder, J. J., Phillips, C. R., & Davis, E. W. *Project Management With CPM, PERT and Precedence Diagramming.* Van Nostrand Reinhold, New York, 1983.

[6] National Aeronautics and Space Administration (NASA). *Analytic Method for Probabilistic Cost and Schedule Risk Analysis.* Retrieved January 31, 2020, from: https://www.nasa.gov/pdf/741989main_Analytic%20Method%20for%20Risk%20Analysis%20-%20Final%20Report.pdf, 2013

[7] Oleinikova, S. A., Kravets, O. J. Approximation of the Distribution Law of the Sum of Random Beta Variables. *International Journal on Information Technologies & Security* , **2** (vol. 9), 2017, pp. 53-64.

[8] The R Project for Statistical Computing: https://www.r-project.org/ Retrieved February 14, 2020,

[9] Vilenkin, S. Y. Determination of Distribution Law of Maximum Time (Critical Path Time). *Avtomatika i Telemekhanika,* **7** (vol. 26), 1965, pp. 1247-1252 (in Russian).

Information about the author:

Deyan Mihaylov – PhD in informatics, chief assistant-professor at University of Economics – Varna, Statistics and Applied Mathematics Department. His area of interest is operations research.

Manuscript received on 17 February 2020