27th IPMA World Congress

The Effects of Different Activity Distributions on Project Duration in PERT Networks

Miklos Hajdua*, Orsolya Bokora

*Szent Istvan University, Ybl Miklos Faculty of Architecture and Civil Engineering, Thokoly ut 74, Budapest, 1146, Hungary

Abstract

The original PERT technique assumes β distribution for the activity durations. The developments of the past decades have partly refuted this concept; and many different distributions have been introduced. In our research, various distributions (uniform, triangular, β) are applied in case of closed large-scale infrastructure projects. In this paper, the possible effects of these distributions on the project duration are investigated. It is shown that the use of different distributions with the same three-point estimation has a smaller effect on the project duration than a 10% difference in the values of the three-point estimation.

© 2014 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of the IPMA.

Keywords: activity distribution; PERT; sensitivity analysis.

1. Introduction

In this paper, we examine the effect of the application of different activity distributions on the distribution of the project duration in PERT networks. We prove – using both artificially created sample projects, and real-life infrastructure projects – that the difference between the results caused by different activity distributions is not significant compared to what a 10% inaccuracy in the estimation of the most likely, optimistic and pessimistic values of the activities can cause to the distribution of the project duration.

The original Program Evaluation and Review Technique (PERT) (Malcolm, et al., 1959) is an activity-on-arrow network with one start and one finish event, which represent the beginning and the end of a project. To accomplish

* Corresponding author. Tel.: +36-1-252-12-70
E-mail address: hajdu.miklos@ybl.szie.hu
the project, certain activities must be carried out according to a given pre-defined sequence. This logic is depicted by a directed, acyclic graph in which the vertices of the graph represent the events, while the arrows represent the tasks to be performed. An event occurs when all preceding activities have been completed; only then can the succeeding tasks start. In this way, the event is used for expressing logical dependencies between activities.

In a PERT network, activity durations are defined by stochastic variables that are assumed to be independent of each other. The distribution of the activity durations follows a so-called PERT-beta distribution. The formula of the beta function is shown below (Eq. 1). In the formula, \( \alpha \) and \( \beta \) are the parameters of the beta distribution; while \( a \) and \( b \) are the endpoints of the domain of \( x \). Outside the interval, \( f(x)=0 \). The distribution is identified as PERT-beta if the \( \alpha \) and \( \beta \) parameters of Eq. (1) are greater than 1 (\( \alpha > 1 \) and \( \beta > 1 \)). This ensures that \( f(x) \) has one maximum, and \( f(x) \) tends to zero at the endpoints of the domain \( f(a)=f(b)=0 \) (Figure 1).

\[
f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}, \quad a < x < b, \quad \alpha, \beta > 0 \quad \text{Eq. (1)}
\]

The mean \( \langle x \rangle \) and variance \( \sigma_x^2 \) of the activity durations in PERT are defined according to Eq. (2) and Eq. (3), respectively:

\[
\langle x \rangle = \frac{a + 4m + b}{6} \quad \text{Eq. (2)}
\]

\[
\sigma_x^2 = \left( \frac{b - a}{6} \right)^2 \quad \text{Eq. (3)}
\]

where \( a, m \) and \( b \) are subjective values, determined by a specialist, representing the optimistic \( (a) \), the most likely \( (m) \) and the pessimistic \( (b) \) durations of the activity (see Figure 1). The process of defining these subjective values is called the PERT three-point estimation method.

The main goal of the PERT analysis is to create the distribution of the project duration. According to PERT theory, the project duration follows a normal distribution, with the mean being the result of time analysis based on activity mean durations \( \langle x \rangle \) and the variance being equal to
\[ \sigma_{PD}^2 = \sum_{x \in CP} \sigma_{x}^2 \quad \text{Eq. (4)} \]

where \( \sigma_{PD}^2 \) is the variance of the distribution of the project duration (PD) and \( x \) represents the activities on the critical path (CP). These calculations are based on the central limit theorem of mathematical statistics.

The theoretical optimistic and pessimistic project durations, that is, the lower and upper bounds of the distribution of the project duration, can be defined as the results of time analysis performed with the optimistic and pessimistic values, respectively.

2. Reviews and developments of PERT

PERT has received a great deal of criticism since its “birth”. The proposed activity distributions, the three-point estimation, the independence of activities, the results of the PERT calculation and the fact that PERT theory omits activity calendars have all been criticized. In the following, only the first aspect is described in detail.

The use of the beta distribution has been criticized by many researchers. During recent decades, researchers have suggested the application of many different distributions other than beta, like the doubly truncated normal distribution (Kotiah and Wallace, 1973), the log-normal distribution (Mohan, et al., 2007), the mixed beta and uniform distribution (Hahn, 2008), the triangular distribution (Johnson, 1997) and the Parkinson distribution (Trietsch, et al., 2012), among others. Some authors – among them Clark (1962) – argue against the introduction of new probability distributions into PERT. According to him: “The author has no information concerning distributions of activity times; in particular, it is not suggested that the beta or any other distribution is appropriate.” In the same line of thinking, Kamburowski (1997) stands by the applicability of the original assumptions (Eq. (2) and Eq. (3)) and opposes those who believe that a different distribution must be introduced. He argues that due to the significant uncertainty and imprecision reflected in the estimates, the precision that we can achieve using Eq. (2) and Eq. (3) is satisfactory. The author strongly agrees with these authors. One of the main purposes of this paper is to justify the above quotation by showing that the usage of different distributions does not result in considerably great differences in the distribution of the project duration, or at least they are smaller than the difference caused by a 10% inaccuracy in the estimation of the most likely values of the activities.

Solutions for determining the distribution of the project duration can be divided into three groups according to the classification by Adlakha (1989) and Elmaghraby (1989), from estimations (Dodin, 1985a) (Kamburowski, 1992) (Yaoi, 2007), through Monte Carlo simulation (Dodin, 1985b) (van Slyke, 1963) (Burt and Garman, 1971), (Sigal, Pritsker and Solberg, 1979) to analytical solutions (Fischer, Saisi and Goldstein, 1985)).

3. The effect of activity distributions on the project duration

3.1. Artificial projects

In this section, two artificially created sample networks are examined. In each case, a Monte Carlo simulation is performed assuming that:

a) all activities follow a PERT beta distribution, which is defined by the original three-point estimation,

b) all activities follow a PERT beta distribution with 10% smaller values for all activity durations (optimistic, most likely, pessimistic) in case of all activities,

c) all activities follow a PERT beta distribution with 10% greater values for all the activity durations (optimistic, most likely, pessimistic) in case of all activities,

d) all activities follow a triangular distribution, using the data defined by the three-point estimation of a)

e) all activities follow a uniform distribution with the optimistic and pessimistic activity times obtained from the three-point estimation
Analyses are performed by ProJack, a general purpose scheduling tool that can perform Monte Carlo simulation. (www.projackmanager.com)

3.1.1. Sample #1

The first – artificial – sample project can be seen on Figure 2. This is a one-chain network, where the same attributes are applied for all the activities in case of a), d) and e), that is a=60 days, m=100 days, b=150 days, 10% smaller values in case of b), and 10% larger values in case of c) for all activities. Activities follow PERT-beta distribution in case of a), b) and c). Activities follow triangular distribution in case of d), and uniform distribution in case of e).

Data for the different cases can be seen in Table 1.

![Figure 2: Sample project #1.](image)

Table 1: Durations and distributions for cases a) – e) of Sample #1 - #2

<table>
<thead>
<tr>
<th>Cases</th>
<th>Optimistic (a)</th>
<th>Most likely (m)</th>
<th>Pessimistic (b)</th>
<th>Distribution Type</th>
<th>Legend of Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case a)</td>
<td>60 days</td>
<td>100 days</td>
<td>150 days</td>
<td>PERT Beta</td>
<td></td>
</tr>
<tr>
<td>Case b)</td>
<td>54 days</td>
<td>90 days</td>
<td>135 days</td>
<td>PERT Beta</td>
<td></td>
</tr>
<tr>
<td>Case c)</td>
<td>66 days</td>
<td>110 days</td>
<td>165 days</td>
<td>PERT Beta</td>
<td></td>
</tr>
<tr>
<td>Case d)</td>
<td>60 days</td>
<td>100 days</td>
<td>150 days</td>
<td>Triangular</td>
<td></td>
</tr>
<tr>
<td>Case e)</td>
<td>60 days</td>
<td>100 days</td>
<td>150 days</td>
<td>Uniform</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 shows the results. It can be seen, that the different activity distributions cause smaller differences in the distribution of the project duration than the difference caused by a +/- 10% percent difference in the three-point estimation.

![Figure 3: Distributions for Sample #1 cases a) –e) ](image)
3.1.2. Sample #2

The second – artificial – sample project can be seen on Figure 4. This network consists of ten parallel paths, where each path is identical to the one-chain network of Sample #1. The same attributes are applied for all the activities in case of a), d) and e), that is $a=60$ days, $m=100$ days, $b=150$ days. In case of b), the values are 10% smaller, while in case of c), they are 10% larger for all activities. Activities follow PERT-beta distribution in case of a), b) and c). Activities follow triangular distribution in case of d), and uniform distribution in case of e).

Data for the different cases can be seen in Table 1.

In order to ensure the requirements for the one start and one finish event, a new start and a new finish event have been included, and paths have been connected to them. (Figure 4)

Results can be seen on Figure 5. Due to multiple critical paths, the distance between cases b) and c) is considerably smaller than the difference between the same cases of Sample #1. Despite this, cumulative distributions derived from triangular and uniform activity distributions remain between the distribution of b) and c). It can be seen that the different activity distributions cause smaller differences in the distribution of the project duration than the difference caused by a +/- 10% percent difference in the three-point estimation. It is interesting to see that uniform distribution results in a more pessimistic distribution than the PERT Beta with the original (60,100,150 days) estimation. This is due to the fact that in case of many parallel paths, the path with the maximum length will define the project duration.

![Figure 4: Sample project #2.](image)

![Figure 5: Distributions for Sample #2 cases a) – e).](image)
3.2. REAL-LIFE PROJECTS

3.2.1. Common Features

Now, after artificially-created networks have been analyzed, two real-life, large-scale infrastructure projects are examined. The greatest difference between the artificial projects and these, real ones is that the latter contained dozens of time constraints. The following time constraints were defined in the sample projects: Must Start On, Must Finish On, Must Start Earlier Than, Must Finish Earlier Than, Must Start Later Than, Must Finish Later Than.

In our research, the Must Start On constraints have been converted to Must Start Later Than constraints. The Must Start Later Than constraint can be considered as a Finish-to-Start relationship between the start activity and the given activity, where the lag time is equal to the difference between the finish of the start task and the start of the given activity expressed in working days. The Must Finish On constraints have been handled the same way, so they have been changed to Must Finish Later Than constraints.

The Must Start Later Than and Must Finish Later Than constraints have remained the same. These could be understood as Finish-to-Start or Finish-to-Finish relationships connecting the start of the project and the given activity, where the lag time is equal to the difference between the finish of the start activity and the start/finish of the given activity expressed in working days.

The Must Start Earlier Than and Must Finish Earlier Than constraints could be substituted with max Finish-to-Start or max Finish-to-Finish relationship defined between the finish of the start activity and the start/finish of the given activity. (Hajdu (1997) wrote about maximal-type relationships in details.) However, the application of these constraints limits the simulation, because the conditions of the network could not be satisfied in those cases where the solution is greater than the given date. Owing to the fact that the aim of the simulation is to get information on the probability of the occurrence of these unfavorable cases, these types of constraints have been deleted.

During the examination of the real-life projects we have followed the same methodology used in the case of the artificial samples. These cases are shown in Table 2. (t stands for the original activity duration used in the Precedence Diagramming network.)

<table>
<thead>
<tr>
<th>Cases</th>
<th>Optimistic (a)</th>
<th>Most likely (m)</th>
<th>Pessimistic (b)</th>
<th>Distribution Type</th>
<th>Legend of Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case a)</td>
<td>m - 40%</td>
<td>t</td>
<td>m + 50%</td>
<td>PERT Beta</td>
<td>[Red]</td>
</tr>
<tr>
<td>Case b)</td>
<td>m - 40%</td>
<td>0.9t</td>
<td>m + 50%</td>
<td>PERT Beta</td>
<td>[Yellow]</td>
</tr>
<tr>
<td>Case c)</td>
<td>m - 40%</td>
<td>1.1t</td>
<td>m + 50%</td>
<td>PERT Beta</td>
<td>[Blue]</td>
</tr>
<tr>
<td>Case d)</td>
<td>m - 40%</td>
<td>t</td>
<td>m + 50%</td>
<td>Triangular</td>
<td>[Purple]</td>
</tr>
<tr>
<td>Case e)</td>
<td>m - 40%</td>
<td>t</td>
<td>m + 50%</td>
<td>Uniform</td>
<td>[Green]</td>
</tr>
<tr>
<td>Case f)</td>
<td>m - 40%</td>
<td>1.15t</td>
<td>m + 50%</td>
<td>PERT Beta</td>
<td>[Black]</td>
</tr>
</tbody>
</table>

Analyses are performed by ProJack, a general purpose scheduling tool that can perform Monte Carlo simulation. (www.projackmanager.com)

3.2.2. Sample #3

This sample project is a highway construction project. The plan is based on the Precedence Diagramming Method. It consists of 1015 activities and 1112 logical relationships. Only minimal-type relationships are defined in the network. There are 552 Finish-to-Start, 239 Start-to-Start, 2 Start-to-Finish and 319 Finish-to-Finish logical dependencies in the plan. 25 milestones and zero hammock activities can be found in the network. The constraints applied in the plan are summarized by Table 3.
Table 3: Constraints of Sample #3

<table>
<thead>
<tr>
<th>Time Constraint</th>
<th>Original Number of Constraints</th>
<th>Number of Constraints after Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Must Start On</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>Must Finish On</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Must Start Earlier Than</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td>Must Finish Earlier Than</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td>Must Start Later Than</td>
<td>41</td>
<td>55</td>
</tr>
<tr>
<td>Must Finish Later Than</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Σ</td>
<td><strong>160</strong></td>
<td><strong>61</strong></td>
</tr>
</tbody>
</table>

According to our rules described earlier, the Must Start On and Must Finish On constraints have been changed to Must Start Later Than and Must Finish Later Than constraints respectively. The Must Finish Earlier Than constraints have been deleted. The 601-day project duration has remained the same after the modifications.

Table 2 summarizes the activity durations and their distributions applied in the six simulations ran for this sample project. Due to the fact that dozens of constraints and five different calendars are defined in the network, the project duration has not changed at the same rate as the activity durations. A 10% decrease in the activity durations has resulted in a mere 3% decrease in the project duration.

![Figure 6: Distributions for Sample #3 cases a) –f)](image)

Results shown in Figure 6 are essentially similar to those of the artificial samples. The function of the uniform distribution goes outside the interval determined by cases b) and c), however, the difference is not considerable. Despite this, case f) is introduced, where a 15% inaccuracy of the PERT three-point estimation is assumed. The green dashed line of case e) stays well within the interval defined by the orange line of case b) and the dark blue line of case f). Therefore our original assumption, according to which the 10% error occurring in the three-point estimation causes greater differences in the distribution of the project duration, could still be considered valid.

3.2.3. Sample #4

This project is also a highway construction project. The project plan is a network based on the Precedence Diagramming Method consisting of 364 tasks and 498 logical dependencies. Only minimal-type relationships are...
defined in the network. There are 245 Finish-to-Start, 176 Start-to-Start, and 77 Finish-to-Finish logical relationships in the plan. 5 milestones and zero hammock activities can be found in the network. The constraints applied in the plan are summarized by Table 4.

Table 4: Constraints of Sample #4

<table>
<thead>
<tr>
<th>Time Constraint</th>
<th>Original Number of Constraints</th>
<th>Number of Constraints after Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Must Start On</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Must Finish On</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Must Start Earlier Than</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Must Finish Earlier Than</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Must Start Later Than</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Must Finish Later Than</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Σ</td>
<td>19</td>
<td>3</td>
</tr>
</tbody>
</table>

The Must Start Earlier Than and the Must Finish Earlier Than constraints have been deleted in accordance with the previously-defined rules. The 295-day project duration has remained the same after the modifications.

Table 2 summarizes the activity durations and their distributions applied in the five simulations ran for this sample project. In case of this sample project, case f) is not dealt with. Due to the fact that dozens of constraints and four different calendars are defined in the network, the project duration has not changed at the same rate as the activity durations. A 10% decrease in the activity durations has caused only 4% decrease in the project duration.

Figure 7 shows the results for this sample project, which are basically the same as those of the artificial projects. The lines indicating cases a), d) and e) stay within the interval defined by cases b) and c). Although the dashed green line of the uniform distribution touches the blue line representing case c), our original assumption is still true. This example also proves that the inaccuracy of the PERT three-point estimation has greater effect on the distribution of the project duration than the type of activity distribution applied. Here the results are not those regular S shaped curves seen in case of the artificial project, and even in case of the previous real-life project. This phenomenon also occurs due to the constraints and calendars defined. Figures 8 show the density functions of case a) (PERT Beta distribution with the original activity durations).

![Figure 7: Distributions for Sample #4 cases a) –e)](image-url)
Recently Hajdu (2013) has shown the dramatic effects of activity calendars on the distribution of the project duration. He has shown using a simple one-chain 10 activity PERT network, with uniform distribution on all the activities, that activity calendars can distort the distribution of the project duration to such an extent which questions the use of the original PERT calculations. Some unusual distributions for project duration can be seen on Figure 9. Minor changes in the calendars have caused the differences between the cases.

4. Conclusions and suggestions

As it is shown in Chapter 2, introduction of different activity distributions plays an important part in the multidirectional development of PERT. However, the practical use of these new distributions has rarely been the subject of a detailed investigation. This issue divides the researchers of this field. Some authors – among them Clark (1962) – argue against the introduction of new probability distributions into PERT, while others apply new
distributions, and try to convince the scientific community that introducing this or that distribution has practical use for planners.

This paper has attempted to add some new aspects to this debate, and investigated this issue from a practical point of view. It has been examined through various sample projects how the application of different activity distributions affects the results compared to the use of the PERT beta distribution. Both the artificial and the real-life projects have shown that +/- 10% difference in the PERT three-point estimation causes greater deviation than the application of different activity distributions. It could be concluded that the usage of different activity distributions does not result in significant differences from a practical point of view. The precision of the three-point estimation plays a much more important role in determining the distribution of the project duration.

The research has to be continued. The results have to be tested in case of other activity durations (e.g. doubly truncated normal distribution, log-normal distribution) and even more practical projects, in order to be able to absolutely confirm our assumption.

References

Hajdu, M.: “Effects of the application of activity calendars on the distribution of project duration in PERT network” in: Automation in Construction, accepted for publication