

# **??**

# **Risk Analysis of Stochastic PERT Graph**

Radek Doskočil, Karel Doubravský

### **Abstract.**

**Purpose of the article.** The paper deals with a time and probability analysis of stochastic graph PERT. The paper focuses on the comparison of two different approaches calculation of probability analysis. Concretely the planning time of the project was calculated. A sample PERT network graph was examined, which comprised 18 nodes and 18 real activities and 6 fictions activities. For the purpose of the analysis, the basic characteristic times were calculated in accordance with traditional approaches related to the PERT method.

**Methodology/methods.** The implementation of the PERT algorithm is based on the critical path method (CPM). It was calculated the basic time charactetistics of the project and identificed the critical path. For probability analysis was also calculated expected value, variance and standard deviance of the activities. For calculation of the planning time was used distribution function of standardized normal distribution. The PERT algorithm is realized by using spreadsheet in the MS Excel.

**Scientific aim.** of the paper is comparison of two different approaches calculation of the probability analysis and their influence on the calculation of the planning time of the project.

**Findings.** Two different approaches calculation of the probability analysis shows on different result of values of project planning time. Approach II better reflects the difference between the values of variances of project activities. The value of variance depends on the input values of three time durations s activity estimates (pessimistic, most likely, optimistic). For higher values of probability there is a bigger difference between the values of planned times that are calculated by two described approaches.

**Conclusions.** The problem was solved using the example project whose model (network graph) contained 18 nodes and 24 activities. For each activity have been known three time estimates (pesimitic, most likely, optimistic). Based on these estimates were calculated expected values of the duration activities and their variances. Expected values of the duration activities were used as input values to calculate the time characteristics. Variances of the activities were used as input values to calculate the variance at the nodes. For these calculations two approaches was used. The expected value of project duration (value of earliest time in last node) was the same for both approaches. For the approach I is a value of the variance in the last node less than for the approach II. These values were used as input data for calculation of planning time of the project at various levels of probability according to the standardized normal distribution. From obtained results dependence between the probability and size of the differences in planned times were observed. This difference increases with a probability going to one. Based on the analysis a recommendation shows to use the approach II under conditions when there are large variations between optimistic (pessimistic) estimates of activity durations and the most likely estimate of activity duration. It causes great differences in values of the variances of the activities. The approach II better reflects this dissimilarity in the variances of the activities. This approach provides longer planning times of the project opposite the approach I.

**Keywords:** network graph, PERT method, node criticality, mean value, probability analysis

**JEL Classification:** C10, C44

### **Introduction**

Project management is nowadays a widely used and discussed discipline. This fact is substantiated by numerous scientific articles, books and publications dealing with these problems (Relich, 2010; Bergantiños, Vidal-Puga, 2009; Garcia, *et all.*, 2005). This discipline is also included in the courses of numerous faculties focusing on economy both in the Czech Republic (Korecký, Trkovský, 2011) and abroad. Experts are also associated in various professional organizations or associations (Společnost pro projektové řízení Česká republika, 2011; International Project Management Association, 2011).

Project managers and other members of the project team use different tools, techniques or methods in project management (CPM, PERT, Gantt diagram, Histogram of sources *etc.*). If we take a closer look at the PERT method, we will find out that in certain cases the manner of their application varies, depending on the approach of individual authors. These differences are based on various methods for calculating the variation of nodes. This fact causes a difference in the results of probabilistic analysis of the project. The scope of probability analysis may vary with each user depending in particular on his/ her needs. The basic ones include computation of the probability of meeting the planned deadlines (Jablonský, 2002), and analysis of node criticality, or activity criticality (Premachandra, 2001).

The paper deals with a PERT method. This method is signed as stochastic method. Its aim is an identification of critical path in a graph. The graph represents a model of a project. The implementation of the PERT algorithm is based on the critical path method (CPM). The paper focuses on the comparison of two different approaches calculation of probability analysis and their influence on the calculation of the planning time of the project. A sample PERT network graph was examined, which comprised 18 nodes and 18 real activities and 6 fictions activities.

### **1. Methods**

The problems under analysis are shown on a PERT network graph comprising 9 nodes and 14 ctivities. Three estimates of activity duration were provided: optimistic, most likely and pessimistic. Subsequently, activity duration mean times (1) were computed according to the following formula (Plevný, Žižka, 2005):

$$
t_{ij} = \frac{a_{ij} + 4m_{ij} + b_{ij}}{6}, \qquad (1)
$$

where:

 $t_{ii}$ 

 $t_{ij}$  activity duration,<br>*a*<sub>*a*</sub> optimistic estima

 $\dot{a}_{ij}$  optimistic estimate of activity duration,<br> $\dot{m}_{ii}$  most likely estimate of activity duration

 $\vec{m}_{ij}$  most likely estimate of activity duration,<br> $\vec{b}_{ij}$  pessimistic estimate of activity duration.

*bij* pessimistic estimate of activity duration.

Variances (2) and standard deviations (3) of activity duration were also calculated. The following formulas were used for their calculation:

$$
\sigma^2 t_{ij} = \frac{(b_{ij} - a_{ij})^2}{36}, \qquad (2)
$$

$$
\sigma t_{ij} = \frac{b_{ij} - a_{ij}}{6} \,. \tag{3}
$$

For the purposes of the time analysis, basic characteristic times were calculated in accordance with traditional approaches. For more detailed information see related publications (Černá, 2008; Wisniewski, 1996).

Using incidence matrix, the earliest times for each node (4) were calculated as follows:

$$
ETN_j = \max\left\{EFT_{ij}\right\},\tag{4}
$$
  
where  $EFT_{ij} = EST_{ij} + t_{ij}$ ,

where:



The latest times for each node (5) were calculated as follows:

$$
LTN_i = \min\left\{LST_{ij}\right\},\tag{5}
$$
\n
$$
\text{where } LST_{ij} = LFT_{ij} - t_{ij},
$$

where:

*LTN<sub>i</sub>* Latest Time of Node,<br>*LFT<sub>ii</sub>* Latest Start Time of A *LFT<sub>ij</sub>* Latest Start Time of Activity,<br>*LFT*<sub>*i*</sub> Latest Finish Time of Activity Latest Finish Time of Activity,  $t_{ii}$ **Activity Duration.** 

The total float of activity (6) from *i*th node to *j*th node was calculated as follows:

$$
TF_{ij} = LTN_j - ETN_i - t_{ij},\tag{6}
$$

where:



There are two generally used approaches for computation of variances needed for the node criticality analysis.

The first approach is based on the fact that the variances of the earliest times of nodes (7) correspond to computed variances of the earliest finish time of activity, out of which that one is selected that appertains to the maximum duration of the earliest finish times of activities incising with the respective node.

$$
\sigma^2 E T N_i = \sigma^2 E F T_{ij} , \qquad (7)
$$
  
for 
$$
\max_i \{ E F T_{ij} \}, j = 1, 2, ..., k ,
$$

where:

*σ2* Variance of the Earliest Time of Node.  $\sigma^2 EFT$ <sub>*ii*</sub> Variance of the Earliest Finish Time of Activity, *k* Number of Nodes.

As regards the first approach, the variances of the latest times for each node (8) correspond to the computed variances of latest start time of activity, out of which that one is selected that appertains to the minimum duration of the latest start times of activities incising with the respective node (Gross, 2003).

$$
\sigma^2 LTN_j = \sigma^2 LST_{ij},
$$
\n
$$
\text{for} \quad \min_j \{LST_{ij}\}, \quad i = 1, 2, \dots, k \,,
$$
\n
$$
(8)
$$

where:

*σ2* Variance of the Latest Time of Node,  $\sigma^2 LST$ <sub>*11</sub>*</sub> Variance of the Latest Start Time of Activity,

*k* Number of Nodes.

The second approach is based on the fact that the variances of the earliest node times (9) correspond to the computed variances of earliest finish time of activity, out of which that one is selected that has the maximum value out of those incising with the respective node.

$$
\sigma^2 ETN_i = \max_i \left\{ \sigma^2 EFT_{ij} \right\}, j = 1, 2, \dots, k , \quad (9)
$$

where:

*σ2* Variance of the Earliest Time of Node, *σ2* Variance of the Earliest Finish Time of Activity, *k* Number of Nodes.

Variances of the latest times for each node (10) correspond to the computed variances of the latest start time of activity, out of which that one is selected that has the maximum value out of those incising with the respective node (Operační analýza [Operational Analysis], 2003):

$$
\sigma^2 LTN_j = \max_j \left\{ \sigma^2 LST_{ij} \right\}, i = 1, 2, \dots, k , \quad (10)
$$

where:

*σ2* Variance of the Latest Time of Node,  $\sigma^2 LST_n$ Variance of the Latest Start Time of Activity, *k* Number of Nodes.

The calculated value (parameter) of the standardized normal distribution (11) will be:

$$
P(T \le PT) = F\left(\frac{PT - ETN_n}{\sigma_{ETN_n}}\right),\tag{11}
$$

where:

*P* Probability*,*

*T* Time,

*PT* Planning Time,

*ETN* Earliest Time of End Node (Time duration of the project),

*σ ETN<sup>n</sup>* Standard Deviation of Earliest Time of End Node.

The sought probability will be the value of the distribution function F(u) (Wonnacott, 1990), which can be found e.g. in the statistical tables, or using appropriate software (Mathews, 2005).

Case study (the PERT algorithm with probabilistic analysis) was calculated by using software MS Excel which provides quickness and comfortable of the solution (Doskočil, Doubravský, 2012). Samuel Bodily (Bodily, 1986) showed the use of spreadsheets for calculation the Operation Research (OR) problems. Spreadsheets provide a simple way how to obtain solutions and how to generate reports. Only few papers have presented a generalized method for implementation of the PERT method in a spreadsheet as the medium (Seal, 2001; Hillier, 2000). Some authors transform this PERT problem to the linear programming and solve it using solver module available in MS Excel. (Ragsdale, 2001).

### **2. Results**

Case study describes a project which includes 18 subactivities. The details of the activities and their predecessor relationships are presented in column 1 and 2 of table 1. The details of the activities and time estimates, i.e. optimistic  $(a_{ij})$ , most likely  $(m_{ij})$ and pessimistic  $(b_{ii})$  of activity duration are shown in column 3, 4 and 5 of Table 1.

|                   |                              | <b>Time of Activity Duration</b> |                      |                             |  |  |  |  |  |  |
|-------------------|------------------------------|----------------------------------|----------------------|-----------------------------|--|--|--|--|--|--|
| <b>Activities</b> | <b>Immediate Predecessor</b> | <b>Optimistic Time (a)</b>       | Most likely Time (m) | <b>Pessimistic Time (b)</b> |  |  |  |  |  |  |
| А                 |                              | 21                               | 24                   | 26                          |  |  |  |  |  |  |
| B                 |                              | 50                               | 56                   | 62                          |  |  |  |  |  |  |
| С                 | A, B                         | 75                               | 80                   | 85                          |  |  |  |  |  |  |
| D                 | С                            | 30                               | 32                   | 34                          |  |  |  |  |  |  |
| E                 |                              |                                  | 4                    | 5                           |  |  |  |  |  |  |
| F                 |                              | 6                                | 8                    | 10                          |  |  |  |  |  |  |
| G                 |                              |                                  |                      |                             |  |  |  |  |  |  |
| H                 | D, E, F, G                   | 2                                |                      |                             |  |  |  |  |  |  |
|                   | H                            | 10                               | 12                   | 14                          |  |  |  |  |  |  |
|                   |                              |                                  | 4                    | 5                           |  |  |  |  |  |  |
| K                 | H                            |                                  | 12                   | 16                          |  |  |  |  |  |  |
| L                 | K                            |                                  | 4                    | 5                           |  |  |  |  |  |  |
| М                 | L                            |                                  | 8                    | 20                          |  |  |  |  |  |  |
| N                 |                              |                                  |                      |                             |  |  |  |  |  |  |
| Ω                 |                              |                                  |                      | 6                           |  |  |  |  |  |  |
| P                 |                              |                                  |                      |                             |  |  |  |  |  |  |
| R                 | O, P                         |                                  |                      | n                           |  |  |  |  |  |  |
| S                 | M, N, R                      |                                  |                      | 6                           |  |  |  |  |  |  |

*Table 1. Representation of the Project.*

*Source: Edited by (Rais, Doskočil, 2011).*





*Source: Own work.*





*Table 3. Times Characteristic of Activities.*

*Source: Own work.*

The graphical representation of predecessor and relationships of project (the network graph) is shown in Figure 1. Network graph consists of 18 nodes,18 real activities and 6 fictions activities.

Table 2 presents the calculated mean time (estimate of expected value) of the activities (column 7) and their variances (column 8). Column 1 to 6 represents informations about the project.

Table 3 contains calculations of the basic characteristic times at the level of activities, including total float, which identifies the critical path. The total float equals zero for the following activities: (1;2), (2;3), (3;4), (4;8), (8;9), (9;10), (10;12), (12;15), (15;17) and (17;18). These activities are therefore critical, and their sequence determines the probable critical path. The estimate of the project duration mean time



*Figure 1. Network graph. Source: Own work.*





*Source: Own work.*





*Source: Own work.*

equals the critical path duration, *i.e.* 200,5 time unit.

Another step in the probability analysis is the calculation of variances of characteristic times at the level of both the nodes (earliest and latest node times) and activities (see Table 3). An incidence matrix is used for practical processing, into which only certain values of characteristic times and their variances are entered. For more detailed information see related publications (Gros, 2003).

The following part of the text summarizes the results of both examined methods of variance computation.

### **Approach I:**

Table 4 represents an incidence matrix of the network graph. Its internal cells contain calculated variances of activity duration (see Table 2). The last column contains variances of the earliest node times calculated using the first approach, i.e. variances corresponding to the calculated activity variances, out of which that one is selected that appertains to the maximum duration of the earliest finish times of activities incising with the respective node. The last line contains variances of the latest node times calculated using the first approach, i.e. variances corresponding to the calculated activity variances, out of which that one is selected that appertains to the minimum duration of the latest start times of activities incising with the respective node.

#### **Approach II:**

Table 5 represents an incidence matrix of the network graph. Its internal cells contain calculated variances of activity duration (see Table 2). The last column contains variances of the earliest node times calculated using the second approach. i.e. variances corresponding to the calculated activity variances. out of which that one is selected that appertains to the maximum variance out of the earliest activity variances incising with the respective node. The last line contains variances of the latest node times calculated using the second approach. i.e. variances corresponding to the calculated activity variances. out of which that one is selected that appertains to the maximum duration of the latest start times of activities incising with the respective node.

#### **3. Discussion**

Table 6 summarizes the results of both approaches of the node criticality computation. The first row defines the node number. The second and third rows give probabilities of their criticality according to the first and second approach.

For calculation of planning time of the project was used (11). Planning times of the project for approach I and approach II computing with probability 0.95 are follows:

Approach I:

$$
P(T \le 205.47) = F\left(\frac{205.47 - 200.5}{3.02}\right) = 0.95.
$$

Approach II:

$$
P(T \le 207.72) = F\left(\frac{207.72 - 200.5}{4.39}\right) = 0.95.
$$

Graphical representations (density function of normal distribution) of previous calculations for approach I (with expeted value 200.5 and standard deviation 3.02) and approach II (with the same expeted value and standard deviation 4.39) are shown in Figure 2.



*Figure 2. Density function of normal distribution for various σ. Source: Own work.*

The values from last column of Table 6 were used as input data for calculation of planning time of the project at various levels of probability according to the standardized normal distribution (Table 7).

From obtained results dependence between the probability and size of the differences in planned times were observed (Table 7). This difference increases with a probability going to one. Based on the analysis a recommendation shows to use the approach II under conditions when there are large variations between optimistic (pessimistic) estimates of activity durations and the most likely estimate of activity duration. It causes great differences in values of the variances of the activities. The approach II better reflects this dissimilarity in the variances of the activities. This approach provides longer planning times of the project opposite the approach I.

*Table 6. Standard deviation of the Earliest Time of Nodes.*

| Number of Node 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 |   |  |  |   |  |  |  |  |  |  |  |  |  |  |  |
|---|---|--|--|---|--|--|--|--|--|--|--|--|--|--|--|
| Approach I  | 3.02 3.02 3.02 3.02 2.97 3.02 2.97 3.02 3.02 3.02 4.39 3.02 4.39 3.00 3.02 3.06 3.02 3.02 |  |  |   |  |  |  |  |  |  |  |  |  |  |  |
| Approach II   |   |  |  | 4.39 4.39 4.39 4.39 4.35 4.39 4.35 4.39 4.39 3.02 4.39 3.02 4.39 3.00 3.02 3.06 4.39 4.39 |  |  |  |  |  |  |  |  |  |  |  |
|   |   |  |  |   |  |  |  |  |  |  |  |  |  |  |  |

*Source: Own work.*





*Source: Own work.*





*Source: Own work.*

In Table 8 probabilities of planning time are showed for both approaches (I and II). In the last row the differences between the values of probability determining by both approaches are calculated.

If the value of difference is less than 0.05 then it can be told that the values of probability determining by approach I and II are equal (can be regarded as the same) from statistical point of view. Otherwise the values of probability are not equal (can be regarded as the different) from statistical point of view. The Table 8 shows that the probabilities determining by both approaches are the same close to the mean time of the project (200.5 time unit) and behind the border of ±3*σ* interval.

# **Conclusion**

The problem was solved using the example project whose model (network graph) contained 18 nodes and 24 activities (18 real. 6 fictions). For each activity have been known three time estimates (pesimitic. most likely. optimistic). Based on these estimates were calculated expected values of the duration activities and their variances. Expected values of the duration activities were used as input values to calculate the time characteristics (Earliest Finish Time of Activity. Earliest Start Time of Activity. Latest Start Time of Activity. Latest Finish Time of Activity and Total float). Variances of the activities were used as input values to calculate the variance at the nodes. For these calculations two approaches was used. The first approach is based on the fact that the variances of the earliest times of nodes corresponds computed variances of the earliest finish time of activity. Out of which that one is selected that appertains to the maximum duration of the earliest finish times of activities incising with the respective node. The second approach is based on the fact that the variances of the earliest node times correspond to the computed variances of earliest finish time of activity. Out of which that one is selected that has the maximum value out of those incising with the respective node. The expected value of project duration (value of earliest time in last node) was the same for both approaches. For the approach I is a value of the variance in the last node less than for the approach II. These values were used as input data for calculation of planning time of the project at various levels of probability according to the standardized normal distribution. From obtained results dependence between the probability and size of the differences in planned times were observed. This difference increases with a probability going to one. Based on the analysis a recommendation shows to use the approach II under conditions when there are large variations between optimistic (pessimistic) estimates of activity durations and the most likely estimate of activity duration. It causes great differences in values of the variances of the activities. The approach II better reflects this dissimilarity in the variances of the activities. This approach provides longer planning times of the project opposite the approach I.

# **Acknowledgments**

This paper was supported by grant FP-S-13-2148 "The Aplication of ICT and Mathematical Methods in Business Management" of the Internal Grant Agency at Brno University of Technology.

### **References**

Bergantiños, G., Vidal-Puga, J. (2009). A Value for PERT Problems. *International Game Theory Review,* 11(4), pp. 419–436.

Bodily, S. (1986). Spreadsheet Modeling as a Stepping Stone. *Interface*, 16(5), pp. 34–52.

Černá, A. (2008). *Metody operačního management (Methods of Operational Management).* Praha: Nakladatelství Oeconomica. 213 pp.

Doskočil, R., Doubravský, K. (2012). Implementation of the PERT Method in MS Excel, In: *Innovation and Sustainable Competitive Advantage*. Istanbul, Turkey: International Business Information Management Association (IBIMA), prepress.

Garcia, P. J. *et al.* (2005). The Two-Sided Power Distribution for the Treatment of the Uncertainty in PERT. *Statistical Methods and Applications*, 14(2), pp. 209–222.

Gros, I. (2003). *Kvantitativní metody v manažerském rozhodování (Quantitative Methods in Management Decisions)*. Praha: Grada Publishing. 432 pp.

Hillier, F. S., Hillier, M. S. (2000). *Introduction to Management Science*. Boston: Irwin-McGraw-Hill, 593 pp.

International project management association (2011). [online] [cit. 2012-04-12] Retrieved from: http://ipma.ch/. Jablonský, J. (2002). *Operační výzkum; kvantitativní modely pro ekonomické rozhodování (Operational Research, Quantitative Models for Economic Decisions).* Praha: Professional Publishing. 323 pp.

Korecký, M., Trkovský, V. (2011). *Management rizik projektu se zaměřením na projekty v průmyslových podnicích (Management of project risks, focusing on projects in industrial companies).* Praha: Grada Publishing, 584 pp.

**Doručeno redakci: 26. 10. 2012 Recenzováno: 5. 6. 2013 Schváleno k publikování: 18. 6. 2013** Mathews, P. (2005). *Design of Experiments with Minitab*. Milwaukee: ASQ Quality Press, 499 pp.

*Operační analýza (Operational Analysis). Podklady k přednáškám a cvičením*. (2003). Plzeň: Fakulta aplikovaných věd, Západočeská univerzita v Plzni.

Plevný, M., Žižka, M. (2005). *Modelování a optimalizace v manažerském rozhodování (Modelling and Optimization in Managerial Decision Making)*. Plzeň: Západočeská univerzita v Plzni, 298 pp.

Premachandra, I. M. (2001). An Approximation of the Activity Duration Distribution in PERT. *Computers & Operations Research*, 28(5), pp. 443–452.

Ragsdale, C. T. (2001). *Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Management Science*. Chula Vista: South-Western College Publishing.

Rais, K., Doskočil, R. (2011). *Operační a systémová analýza I (Operational and System Analysis I).* Brno: VUT v Brně, Fakulta podnikatelská, 125 pp.

Relich, M. (2010) *Assessment of task duration in investment projects.* Management, 14(2), pp. 136–147.

Seal, K. C. (2001). A Generalized PERT/CPM Implementation in a Spreadsheet. *INFORMS Transaction of Education*, 2(1), pp. 16–26.

Společnost pro projektové řízení Česká republika (Project Management Company Czech Republic). (2011). [online] [cit. 2012-03-15] Retrieved from: http://www.cspr.cz/.

Wisniewski, M. (1996). *Metody manažerského rozhodování (Methods of Managerial Decisions)*. Praha: Grada Publishing, 512 pp.

Wonnacott, T. H., Wonnacott, R. J. (1990). *Introductory Statistics for Business and Economics*. Oxford: Wiley, 416 pp.

#### **Ing. Radek Doskočil. Ph.D.**

Vysoké učení technické v Brně. Fakulta podnikatelská. Ústav informatiky. Kolejní 2906/4. Brno. Czech Republic tel.: 54 114 3722. e-mail: doskocil@fbm.vutbr.cz

#### **Ing. Karel Doubravský. Ph.D.**

Vysoké učení technické v Brně. Fakulta podnikatelská. Ústav informatiky. Kolejní 2906/4. Brno. Czech Republic tel.: 54 114 3718. e-mail: doubravsky@fbm.vutbr.cz