CRITICAL PATH VALUE ESTIMATION METHOD TAKING INTO ACCOUNT THE SKEWNESS OF PERT DISTRIBUTIONS

Małgorzata KOWALCZYK, Bożena ZWOLIŃSKA

AGH University, Faculty of Mechanical Engineering and Robotics, Cracow, Poland, EU, bzwol@agh.edu.pl

Abstract

The more complex a production process is, the more difficult it is to estimate accurately the production tasks times. The Project Evaluation and Review Technique (PERT) is one of the most frequently used methods for evaluating the project lead times (also sets of production tasks). Various variants of the method are used depending on the estimation accuracy degree and available computing power. The basic method for determination of the critical path duration is called the classical PERT method. One of its best advantages is a relatively low-level calculation complexity. The disadvantage, on the other hand, is a relatively low estimation accuracy resulting from many generalizations. The literature on the subject matter describes two group of algorithms used to improve the critical path duration estimates relative to the classical PERT method. The first group refers to the estimation improvement of variance and standard deviation of random variables that define the tasks times. The other group focuses on the fit improvement of the probability distribution of random variable. Each method has its advantages resulting from the better accuracy in comparison with the classical PERT, but the calculation complexity grows along with the increasing estimation accuracy.

The paper aims at developing the estimation method for the critical path duration based on a modified PERT method. The presented method of determining the cumulative execution time of the sets of tasks accounts for skewness of empirical distributions which frequently occurs in the processes. The studies verifying the method's soundness have been made. The results have been compared with the values obtained with the use of the classical method. Then, the lead times have been determined with the use of the developed method for a real production facility.

Keywords: PERT and beta-PERT distribution, critical path

1. INTRODUCTION

In the analysis of complex production systems, the most frequently used graph theory models are Critical Path Method (CPM) and Project Evaluation and Review Technique (PERT). The classical network analyses methods cover the time planning area. In business units, the economic aspect is equally important, so CPM-COST and PERT-COST are often used to consider the technical possibilities of project time shortening with simultaneous tasks cost reduction [1]. These four network types allow examining the project of undetermined logical structure. The PERT and PERT-COST methods account for probabilistic character of individual events (tasks) by using PERT or beta-PERT. The final result is a single-variant determination of network of relationships between individual tasks. The earliest and latest times of individual events and time reserve between individual processing stages are determined in each case [2]. These methods are helpful in determining the schedule of individual tasks in a project comprising many stages independent of each other. The main goal is to establish the critical path which determines the cumulative project lead time. In case of complex production systems, in which at a given Δt there are two critical paths independent of each other, it is difficult to determine the total project lead time because it is absolutely dependent on random time values in individual routes. An indisputable advantage of these methods is that they can be used for systems parallel in series in which technologically identical production times may have different durations. One of the main limitations are complex algorithms for determination of cumulative time values for cases in which we wish to obtain a high accuracy of model calculations relative to real systems. Using the classical methods with lower calculation complexity yields correspondingly less accurate results. This is sufficient in many cases. There are, however, business areas where the network analyses should be used because of their advantages, but the classical method is not sufficiently accurate due to numerous generalizations. Therefore, the literature on the network analyses includes many solutions to improve the weaknesses. There are two basic groups: 1) evaluate the variance or standard deviation of individual tasks more accurately; and 2) evaluate the probability in individual nodes more accurately [3]. Each attempt to improve the accuracy of network analyses relative to the classical approach is connected with the increase of calculations complexity. This in turn makes the improved method unsuitable for cases where the precise accuracy is not required. The research on improving the evaluation accuracy of cumulative critical paths times is continuing.

The paper intends to develop an algorithm for determining the cumulative project duration according to the PERT method taking into account various skewness values of individual tasks. The developed algorithm is an attempt to modify the classical PERT method which scales the expected value of a single task using the skewness contribution parameter. The skewness contribution parameter is determined based on the percent share of skewness of an individual distribution relative to all skewness values of the critical path. The proposed method does not increase significantly the level of complication and calculation complexity relative to the classical PERT method, but is its more accurate alternative.

2. DESCRIPTION OF THE PERT METHOD AND PERT AND BETA-PERT DISTRIBUTIONS

The PERT and beta-PERT distributions belong to the family of continuous probability distributions and are a transformation of the beta distribution. The transformation involves rescaling the random variable carrier from interval *[0,1]* to interval *[a,c]* when: $a<c$ [4]. Value *a* is the minimum and *c* is the maximum of the random variable distribution. The PERT and beta-PERT distributions also have the shape parameter *k*. In the PERT distribution, the parameter is precisely defined - it always equals four ($k=4$). Consequently, the PERT distribution is defined by three parameters: *a* - minimum distribution value, *b* - modal distribution value, *c* - maximum distribution value. And the beta-PERT is defined by four parameters: *a, b, c, k*, for $k>4$. Then, the expected value of random variable conforming to the PERT distribution is the weighted mean of the minimum and maximum value and with the quadruple weight of the most probable value [5]. Hence:

$$Et_i = \frac{a + 4 \cdot b + c}{6} \tag{1}$$

On the other hand, in the beta-PERT distribution the expected value is determined according to the formula:

$$Et_i = \frac{a + k \cdot b + c}{k + 2} \tag{2}$$

where: Et_i - expected value of random variable t_i conforming to the PERT distribution; *a* - minimum distribution value, *c* - maximum distribution value, *b* - modal distribution value of random variable t_i, *k*, - modal value multiplicity (for PERT always $k=4$).

Many literature items use the approximations with the PERT distribution for empirical data that have more than four multiplicities. This is a significant simplification, however in majority of cases is considered sufficiently accurate. Note, however, that the use of the PERT distribution is more accurate when the multiplicity of occurrence of the most probable value equals 4. A modification of the PERT distribution with any modal value multiplicity greater than four has been applied in the beta-PERT distribution, allowing the tasks durations to be determined with a higher accuracy. However, due to the level of difficulty and complex calculations, the beta-PERT distribution is not widely used. **Figure 1** presents the distribution comparison diagrams: PERT with triangular, and beta-PERT with triangular.

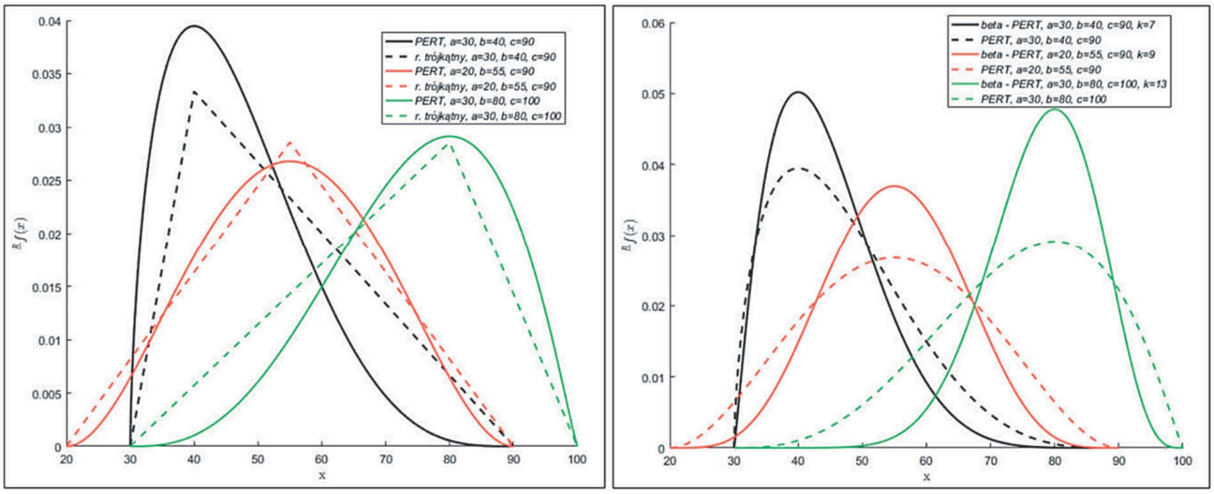

Figure 1 Distribution comparison: PERT with triangle and beta-PERT with PERT

There are three often used methods of matching (estimation) of distribution parameters: *a, b, c* and *k* for a simple sample (e.g. set of production tasks completion times determined by timekeeping:

1) reading the values directly from the dataset, i.e.: *a=min, c=max*, etc.;
2) determining the parameters $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{k}$ using the method of moments;
3) determining the parameters $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{k}$ using the maximum likelihood method.

Various estimation methods for distribution parameters are used depending on the required degree of accuracy and available computing power.

The estimation of determination of optimum parameters $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{k}$ for the cases discussed in the paper has been carried out using the maximum likelihood method. Finally, after many rearrangements, we obtained:

- optimal parameters for the PERT distribution:

$$\begin{cases} \hat{a} = \bar{t} + 2A_s \cdot s - s \cdot \sqrt{4 \cdot A_s^{\ 2} + 7} \\ \hat{b} = \bar{t} - A_s \cdot s \\ \hat{c} = \bar{t} + 2A_s \cdot s + s \cdot \sqrt{4 \cdot A_s^{\ 2} + 7} \end{cases} \tag{3}$$

- optimal parameters for the beta-PERT distribution:

$$\begin{cases} \hat{a} = \bar{t} + 2A_s \cdot s - s \cdot \sqrt{4 \cdot A_s^{\ 2} + \left(\hat{k} + 3\right)} \\ \hat{b} = \dfrac{\bar{t} \cdot \hat{k}}{4} - A_s \cdot s \\ \hat{c} = \bar{t} + 2A_s \cdot s + s \cdot \sqrt{4 \cdot A_s^{\ 2} + \left(\hat{k} + 3\right)} \\ \hat{k} = \dfrac{96A_s^{\ 2} - 63K_s + 147 + \sqrt{64A_s^{\ 2}\left(144A_s^{\ 2} - 63\left(K_s - 1\right)\right) + 343K_s + 441}}{14\left(K_s - 3\right)} \end{cases} \tag{4}$$

where: α and β - distribution parameters determined by formulas [6]:

$$\alpha = 1 + k\left(\frac{b-a}{c-a}\right) \tag{5}$$

$$\beta = 1 + k\left(\frac{c-b}{c-a}\right) \tag{6}$$

and

- average from the sample determined according to: $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ (7)
- standard deviation from the sample determined according to: $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2}$ (8)
- skewness from the sample determined according to: $A_s = \dfrac{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^3}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2}^{\,3}}$ (9)
- kurtosis from the sample determined according to: $K_s = \dfrac{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^4}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2}^{\,4}} - 3$ (10)

The necessary condition in case of using the PERT and beta-PERT distributions to describe the task durations on a single machine is that the following assumptions are met:

- ratio of skewness from the empirical sample must be in the interval *:

for PERT distribution:

$$A_s \in \left[-\sqrt{\frac{7}{5}};\sqrt{\frac{7}{5}}\right] \tag{11}$$

for beta-PERT distribution:

$$A_s \in \left[\frac{\frac{3(k-4)}{2}\cdot\frac{\bar{t}}{s} - \sqrt{(4-k)^2\cdot\left(\frac{\bar{t}}{s}\right)^2 + 20(k+3)}}{10};\frac{\frac{3(k-4)}{2}\cdot\frac{\bar{t}}{s} + \sqrt{(4-k)^2\cdot\left(\frac{\bar{t}}{s}\right)^2 + 20(k+3)}}{10}\right] \tag{12}$$

* determined skewness interval limits result from the distribution properties;

- relationship between the distribution parameters: $0 < a < b < c$; (13)
- kurtosis from the empirical sample $K_S > 0$; (14)
- number of manufactured semi-finished products: $n = [6,30]$. (15)

It is possible to use a distribution for the sample size greater than 30, however, due to the calculation complexity, in practice e.g. a Gauss distribution is used when the conditions for its application are met [7].

The PERT method involves the time estimation according to the expected value of the PERT distribution as per the formula (1) (either (2) for greater accuracy of calculations). The variance is determined in accordance with the formula:

$$D^2 t_i = \left(\frac{c-a}{6} \right)^2 \tag{16}$$

where: $D^2 t_i$ - variance of a single random variable t_i; a and c - estimated minimum and maximum values of the PERT distribution, respectively.

Knowing that the times of individual production tasks are independent of one another, the standard deviation of cumulative production times will be calculated according to the classical PERT method. Hence:

$$DT = \sqrt{\sum_{i=1}^{n} \left(\frac{c_i - a_i}{6} \right)^2} \tag{17}$$

where: DT - standard deviation of random variable T; T - random variable defining the cumulative duration of all tasks on the critical path; $\hat{a}_i$ and $\hat{c}_i$ - estimated minimum and maximum values of PERT distribution of a single random variable t_i for $i = 1,2,...,n$, where n is the maximum number of tasks on the critical path.

Using the linearity property of the expected value of independent random variables t_i for $i = 1,2,...,n$, the ET can be calculated according to the formula:

$$ET = \sum_{i=1}^{n} E(t_i) \tag{18}$$

where: ET - expected value of random variable T, which defines the cumulative value of the most probable duration of production tasks on the critical path (ET defines the lead time of an order or production, depending on the assumptions).

3. ALGORITHM FOR EXPECTED VALUE DETERMINATION ACCOUNTING FOR SKEWNESS

In majority of cases, the histograms of production tasks durations are right-skewed distributions. In such distributions, a significant portion of the probability mass is concentrated on the left-hand side of the plot. The classical PERT method does not account for this aspect. Hence, the proposition to develop an algorithm for estimating the cumulative value of the expected duration of the critical path tasks which will include the influences of the existing skewness values. The developed algorithm consists of five calculation steps. Before the calculations, the data on production tasks durations are collected by means of timekeeping. Hence, for the resultant simple sample of independent random variables t_i, for $i = 1,2,...,n$, where: t_i - random variable defining the task duration on the i th production facility; n - number of facilities performing the production processes on the critical path. Then:

$$\forall\, i = 1,2,...,n \quad t_i \sim F_i \tag{19}$$

where each random variable t_i is defined by the distribution of the cumulative distribution function F_i.

The approximation of production times is correct if $\forall\ i = 1,2,...,n$ the following conditions are met:

- the distribution carrier is limited with the values: minimum - a_i, maximum - c_i; in addition: $a_i < c_i$;
- for the simple sample of distribution t_i the mode multiplicity $b_i = 4$ and $a_i < b_i < c_i$.

When these conditions are met, $\forall\ i = 1,2,...,n$ the random variable t_i can be approximated using the PERT distributions with parameters $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$ - system of equations (3).

The calculation methodology according to the proposed algorithm (five successive steps) is presented below.

Step 1 - determine the expected value $E(t_i)$ of variable t_i defining the task duration on a single facility (machine).

$\forall i = 1,2,...,n$:

$$E(t_i) = \frac{\hat{a}_i + 4\hat{b}_i + \hat{c}_i}{6} \tag{20}$$

where: $E(t_i)$ - expected value of random variable t_i, $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$ - optimal values, respectively: minimum, modal and maximum.

Step 2 - determine the skewness of random variable t_i

$\forall i = 1,2,...,n$:

$$A_S(t_i) = \frac{2(\beta_i - \alpha_i)\sqrt{\alpha_i + \beta_i + 1}}{(\alpha_i + \beta_i + 2)\sqrt{\alpha_i + \beta_i}} \tag{21}$$

where: $A_S(t_i)$ - skewness of random variable t_i; α_i, β_i - shape and scale parameters of the probability density function of random variable t_i determined according to the formulas: (5) and (6).

Step 3 - determine the weights $W_{A_S(t_i)}$ of proportional skewness contribution $A_S(t_i)$

$\forall i = 1,2,...,n$ such that $A_S(t_i) \neq 0$ determined is the weight $W_{A_S(t_i)}$ of skewness $A_S(t_i)$ according to the formula:

$$W_{A_S(t_i)} = \frac{|A_S(t_i)|}{\sum_{i=1}^{n} |A_S(t_i)|} \tag{22}$$

where: $W_{A_S(t_i)}$ - weight of skewness contribution for random variable t_i; $A_S(t_i)$ - skewness of variable t_i.

The weights $W_{A_S(t_i)}$ are determined only for non-zero $A_S(t_i)$, i.e. the distribution of random variable t_i is asymmetrical (left- or right-skewed).

Step 4 - determine the parameter $P_{A_S(t_i)}$ scaling the contribution of expected value $E(t_i)$

$\forall i = 1,2,...,n$ such that $A_S(t_i) \neq 0$ determined is the scaling parameter $P_{A_S(t_i)}$ of expected value $E(t_i)$ according to the formula:

$$P_{A_S(t_i)} = \frac{W_{A_S(t_i)} \cdot 1}{\frac{1}{m}} = W_{A_S(t_i)} \cdot m \tag{23}$$

where: $m = n - l$ for l defining the number of random variables whose $A_S(t_i) = 0$; $P_{A_S(t_i)}$ scaling parameter for contribution of expected value $E(t_i)$.

If for a facility i, skewness $A_S(t_i) = 0$, then $P_{A_S(t_i)} = 1$

Step 5 - determine the cumulative expected value ET for the critical path

In the classical PERT method, for the determination of the cumulative duration of tasks on the critical path - formula (18) it was assumed that the number of individual t_i is so great that the central limit theorem of the sum of mean production tasks times can be used. This assumption was used in the algorithm proposed by the authors, and in addition the skewness of individual variables t_i was accounted for. In formula (18), the parameters scaling the expected value $E(t_i)$ are equal to one (that is $P_{A_S(t_i)} = 1$). In the proposed method, the cumulative expected value of the critical path is determined according to:

$$ET = \sum_{i=1}^{n} P_{A_S(t_i)} \cdot E(t_i) \tag{24}$$

where: ET - cumulative value of the critical path; $P_{A_S(t_i)}$ - scaling parameter for contribution of expected value $E(t_i)$.

The classical PERT method is based on assumptions that are not always met. The knowledge of exact distribution of task times in the network of events allows avoiding many errors and estimating the task times with a better accuracy. The application of more accurate PERT methods causes some calculation difficulties resulting from the relationships between random variables.

4. CONCLUSION

The paper presents an algorithm for determination of the expected value of cumulative production task times based on the modified classical PERT method. The proposed algorithm is used to determine ET taking into account the scaling parameters $P_{A_S(t_i)}$ of empirical skewness of individual random variables t_i. The scaling parameters $P_{A_S(t_i)}$ were determined taking into account $W_{A_S(t_i)}$ that is the weight contribution of skewness $A_S(t_i)$ of individual $t_i \sim PERT(a_i, b_i, c_i)$.

The purpose of the developed algorithm is to improve the estimation of the expected critical path lead time. Please note that the time estimate obtained with the use of PERT method is correct when the network contains only one dominating critical path. When the dominating critical path is not there, the task completion times are underestimated. The proposed algorithm should be subject to a detailed validation in order to verify its effectiveness in practice. The verification of correctness should account for the demand seasonality and variability as well as for the product range variations that occur in the studied class of production systems. The obtained results should form a basis to develop a tasks scheduling model for the studied class of convergent production systems.

REFERENCES

[1] COOPER ORDONEZ, Robert Eduardo, VANHOUCKE, Mario, COELHO, Jose, ANHOLON, Rosley and NOVASKI, Olivio. A Study of the Critical Chain Project Management Method Applied to a Multiproject System. *Project Management Journal*. 2019. vol. 50, no. 3, pp. 1-13.

[2] VANHOUCKE, Mario. Project management with dynamic scheduling- Baseline scheduling, risk analysis and project control. *Springer-Verlag*. Berlin, Heidelberg, Germany 2013.

[3] MILIAN, Zdzisław. Methods for determining the distribution of time in the implementation of construction projects in acyclic stochastic networks. *Wydawnictwo Politechniki Krakowskiej*. Kraków 2006.

[4] DEVORE, Jay L. Probability and Statistics for Engineering and the Sciences. *Brooks Cole Pub Co*. Boston 2015.

[5] ERDEI, Edina, POPP, József and OLAH, Judit. Comparison of time-oriented methods to check manufacturing activities and an examination of their efficiency. *LogForum*. 2018. vol. 14, no. 3, pp. 371-386.

[6] PLEGUEZUELO, Rafael Herrerias, PEREZ, Jose Garcia and RAMBAUD, Salvador Cruz. A note on the reasonableness of PERT hypotheses. Operations Research Letters. 2003. vol. 31, pp. 60-62.

[7] ZWOLIŃSKA, Bożena. Modeling convergent processes in complex production systems. *AGH University of Science and Technology Press*. Krakow 2019.