PAPER • OPEN ACCESS

Features of a numerical characteristics estimation for the execution time of complex series-parallel operations

To cite this article: S A Oleinikova *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 042005

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Features of a numerical characteristics estimation for the execution time of complex series-parallel operations

S A Oleinikova[1], E V Bolnokina[2] and O Ja Kravets[1,3]

[1] Voronezh State Technical University, Voronezh, Russia
[2] NIIAA, Moscow, Russia

[3] E-mail: csit@bk.ru

Abstract. The object of research in this paper is stochastic multistage systems, the service process in which a set of series-parallel operations with a random duration of service is performed. The subject of the study is the random variable which is the service time of the entire application. The aim is to evaluate the analytical capabilities to accurately assess the characteristics of the studied value. The possibility of analytically obtaining the characteristics of a random variable is investigated, as a result of which a conclusion is made about the usefulness of connecting numerical apparatus. Thus, the features of estimating the numerical characteristics of a random variable describing the duration of a complex of series-parallel operations are analyzed.

1. Introduction

The process of functioning of complex service systems for the execution of a complex series-parallel operation for the maintenance of the application is investigated. The execution time of each individual operation is a random variable. It is necessary to estimate the duration of service of the entire application. This task belongs to a class of network planning and management (project management) tasks. For tasks with a random duration of work, the PERT (Program Evaluation and Review Technic) method has been developed, which allows not only the execution time of individual operations to be estimated, but also the duration of maintenance of the entire application [1, 2, 3].

However, studies have shown that estimations of the method do not always take place in real conditions. In this regard, PERT formulas will also have some error. To minimize this error, it is necessary to develop a new analytical tool that allows the numerical characteristics of the desired random variable to be estimated with the highest possible level of accuracy.

This paper describes the problem statement; the inaccuracy of estimates of the PERT method is shown experimentally, and this will increase in the presence of several critical paths. The following is a description of the duration of the service, which is the maximum of random variables, each of which represents the total time of work and is set in a certain way. Having described the numerical characteristics of this random variable, we analysed the possibilities for analytically finding the mathematical expectation and variance.

2. Statement of the problem and its features

A multi-stage system, the input of which receives a stream of applications, is considered. Each application requires the implementation of a set of mutually dependent (series-parallel) operations for

its service. Each operation is defined by its duration, which is a random variable, and mutual dependence (a set of previous or subsequent operations). It is required to estimate the numerical characteristics (expectation and variance) of a random variable describing the duration of service of the whole complex of operations.

Let us analyse the features of this problem. In the deterministic case where the execution time of each task is known in advance and unchanged, the value of the maintenance duration can be estimated as follows:

$$T_{fact} = \sum_{k:w_k \in K} t_k \tag{1}$$

Here K is the set of operations forming the critical path; w_k is the regular operation belonging to the critical path; t_k is the duration of this operation.

In the stochastic case, there is uncertainty associated with the possibility of changing the critical path in the maintenance process. Let us explain this with a simple example. Consider the network graph shown in figure 1.
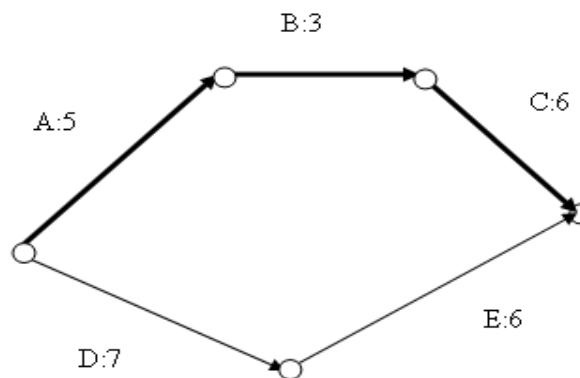

Figure 1. The example of the network graph.

In this figure, the letters denote the operations, and the numbers denote their approximated duration (estimate of mathematical expectation). As can be seen from this figure, the critical path is determined by the operations A, B, and C. However, if, for example, operation C is completed two units earlier, then operations D and E will be critical. Thus, it is impossible to determine in advance either the critical path or the set of operations lying on it.

In addition, the specificity of the duration of operations and their mutual dependence can be such that the problem can have two critical paths. In this case, there are some difficulties associated with estimating the numerical characteristics of the desired random variable.

Let us analyse the most common method of solving this problem, PERT. This method is intended for the formation of service schedules in stochastic systems in interdependent work. Let $\xi_1,\ldots,\xi_k$ be the random variables describing the operations that stand on the critical path. Although all the operations in the system are dependent, it is assumed that the duration of any operation does not affect the duration of other operations. Hence the random variables $\xi_1,\ldots,\xi_k$ are independent. The calculation formulas of PERT for the expectation and variance of the desired duration are as follows:

$$M(\xi) = M(\xi_1) + \ldots + M(\xi_k), \tag{2}$$

$$D(\xi) = D(\xi_1) + \ldots + D(\xi_k). \tag{3}$$

As noted earlier, there may be cases where the network graph will contain two or more critical paths. In this case, the formula (2) will remain suitable for calculating the mathematical expectation (since all critical paths will have the same mathematical expectation). However, the formula for

variance will change. In [1], the variance is estimated as the maximum of all variances, each of which will be calculated for its critical path:

$$D(\xi) = \max(D(\eta_1), \dots, D(\eta_n)) \ . \tag{4}$$

Here ξ is the desired value; n is the number of critical paths; $\eta_1, \dots, \eta_n$ are random variables representing the total duration of all operations in the path $1, \dots n$ respectively.

3. Experimental estimates of service duration

Let us conduct an experiment, the purpose of which is to estimate the duration of service if it is necessary to perform a set of series-parallel operations. At the start of the experiment, the network graph, which is a set of mutually dependent operations, will be submitted. It is assumed that the duration of each operation is distributed according to the law of beta [4]. The rationale for this assumption is presented in [5]. The aim of the experiment is to obtain a sample, each element of which is the duration of the service application (when drawing all random variables describing the duration of individual operations).

Based on this sample, numerical characteristics such as expectation (sample mean) and variance (sample variance) are estimated. The obtained results are compared with the expected analytical results calculated by formulas (2) and (3) (or (4) with equal variances).

The experiment includes many independent experiments, in each of which:

- the random variables distributed according to the beta law are simulated, the aim of which is to simulate the performance of operations;
- the Xi value is determined, which is the service time of the entire application (the moment of completion of the last operation, taking into account their mutual dependence).

At the end of the test series, the sample mean and the sample variance of random variables Xi are calculated.

Consider the simulation of random variables distributed according to the law of beta in detail. The input of the corresponding subroutine receives the expectation and variance of a random variable. Firstly, the parameters p and q of the beta distribution are determined. The following formulas were obtained in [6]:

$$p = \frac{(b - M(\xi)) \cdot (M(\xi) - a)^2 - D(\xi) \cdot (M(\xi) - a)}{D(\xi) \cdot (b - a)} \tag{5}$$

and

$$q = \frac{(b - M(\xi))^2 \cdot (M(\xi) - a) - D(\xi) \cdot (b - M(\xi))}{D(\xi) \cdot (b - a)} \ . \tag{6}$$

The second stage consists in obtaining a random variable with the parameters (5) and (6).

Provided below, without loss of generality, is a small fragment of the results for the trivial case when the project maintenance is given by a set of two parallel operations. In the case where the expected value of these operations is the same, the results are as shown in table 1. The table presents the results for random variables distributed according to the beta law with expectation M and the variances D_1 and D_2, respectively. Based on formulas (2) and (4), the expectation and variance of the service time of requests are M^*_p and D^*_p. Similar characteristics estimated using sample values are M^*_s and D^*_s.

Table 1. A fragment of the results of the experiment for
two random variables.

Initial data			Planned characteristics		The actual characteristics	
M	D_1	D_2	M^*_p	D^*_p	M^*_s	D^*_s
15	2,97	3,78	15	3,78	16,05	2,449
12	0,33	0,22	12	0,33	12,33	0,23
13	0,75	0,5	13	0,75	13,44	0,432
16	3	2	16	3	16,90	1,795
16	3	2	16	3	16,90	1,767

Based on the values from this table, it is obvious that the actual mathematical expectation turned out to be significantly higher than planned, while the variance was lower.

Let us consider the result of three parallel operations. Suppose that one of the three random variables has the expected value of 5.5, and the other two – 5. According to the PERT method, the expected value of the resulting random variable is 5.5. Let us present a fragment of the results of experiments, which differ in terms of the range of random variables (table. 2).

Table 2. A fragment of the results of the experiment
for the three random variables.

Planned characteristics		The actual characteristics	
M^*_p	D^*_p	M^*_s	D^*_s
5,5	0,6944	5,9454	0,6216
5,5	0,4444	5,7971	0,2345
5,5	1	6,0323	0,3399
5,5	1,7778	6,3417	0,9126
5,5	1	6,0089	0,472
5,5	0,6944	5,7223	0,2309
5,5	1	5,8526	0,4434
5,5	1	5,8805	0,2962
5,5	1,3611	6,0444	0,4638
5,5	1	5,9636	0,5404

We will conduct an experiment in similar conditions except for the values of the mathematical expectations of random variables. Let now two random variables have mathematical expectation 5.5, and one – 5. This will be similar to the case where the network graph has two critical paths. The results are shown in table 3.

Table 3. A fragment of the results of the experiment
for the three random variables.

Planned characteristics		The actual characteristics	
M^*_p	D^*_p	M^*_s	D^*_s
5,5	0,6944	6,113	0,5553
5,5	0,4444	5,9484	0,2081
5,5	1	6,177	0,3074
5,5	1,7778	6,5067	0,8757
5,5	1	6,1602	0,3931
5,5	0,6944	5,8768	0,3048
5,5	1	6,0245	0,3709

5,5	1	6,0634	0,3505
5,5	1,3611	6,2112	0,4066
5,5	1	6,1143	0,4878

Compared to the previous case, it can be seen that the actual expectation is increased, and the variance is reduced.

Thus, on the basis of experimental data, fragments of which are presented in tables 1, 2 and 3, the following conclusions can be drawn:

- the evaluation of the duration of the service request, proposed by the PERT method, is lower than the actual duration of service;
- the variance of the random variable describing the service duration, proposed by the PERT method, is higher than the actual variance;
- by increasing the number of critical paths, the expected expectation of a random variable describing the duration increases (as opposed to the PERT method);
- by increasing the number of critical paths, the expected variance of the random variable describing the duration decreases (as opposed to the PERT method).

Thus, it is necessary to develop a mathematical apparatus to obtain more accurate estimates compared to the PERT method.

4. Analytical description of the random variable under study

In this part, an analytical description of a random variable representing the service duration for stochastic systems will be given. To do this, consider all the paths of the network graph (i.e. all the chains that allow you to move from the initial to the final state). Each path will be determined by a variety of operations. Let ξ_i, i=1,…,n be a random variable that determines the total duration of all work that stands in the way of i. Then the required random variable ξ can be described as follows:

$$\xi = \max(\xi_1,...,\xi_n). \tag{7}$$

Thus, the problem can be formulated as follows. Let there be many random variables. It is necessary to find the expectation and the variance of the random variable, which is given by the formula (7).

Similar problems are considered in [7, 8, 9]. However, the principal difference of the studied problems, which relate to the theory of extreme values, is the identity of random variables included in the formula (5). The case where each individual random variable has its own distribution interval and its own parameters has not yet been investigated.

Consider a particular problem. Let

$$\eta = \max(\xi_1, \xi_2). \tag{8}$$

Random variables are distributed in the intervals $[a_1, b_1]$ and $[a_2, b_2]$, respectively. It is necessary to find the expectation and variance of a random variable (8). According to the formulas of probability theory, we obtain [7]:

$$M(\eta) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \max(x, y) f_{\xi_1}(x) f_{\xi_2}(y) dx dy \tag{9}$$

and

$$D(\eta) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} (\max(x, y))^2 f_{\xi_1}(x) f_{\xi_2}(y) dx dy - (M(\eta))^2. \tag{10}$$

It is known that the intervals $[a_1, b_1]$ and $[a_2, b_2]$ are overlapping, i.e. $b_2 > a_1$ and $b_1 > a_2$. Without loss of generality, suppose that $a_2 < a_1$. Then:

$$M(\eta) = \int\limits_{a_1}^{b_1} \int\limits_{a_2}^{b_2} \max(x, y) f_{\xi_1}(x) f_{\xi_2}(y) dx dy = \int\limits_{a_1}^{b_1} f_{\xi_1}(x) \left(\int\limits_{a_2}^{a_1} \max(x, y) f_{\xi_2}(y) dy + \right.$$
$$\left. + \int\limits_{a_1}^{b_2} \max(x, y) f_{\xi_2}(y) \right) dy dx \tag{11}$$

Next, it is necessary to define the maximum function:

$$\max(x, y) = \begin{cases} x, x \geq y \\ y, x < y \end{cases} \tag{12}$$

and divide the integral into components depending on which of the arguments is greater.

Let us proceed to the definition of densities in formulas (9) and (10). Since each random variable ξ_i, $i=1,2$, is a set of consecutive operations, it will be determined by the formula:

$$\xi_i = \xi_{i1} + \ldots + \xi_{ik}. \tag{13}$$

In [13] it was shown that the sum of beta values can be approximated by the beta value. In particular, if random variables $\xi_1, \ldots, \xi_n$ have a beta distribution in the intervals $[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n]$ with parameters $(p_1, q_1), (p_2, q_2), \ldots, (p_n, q_n)$ respectively, then the final density will have the form:

$$f_\xi(x) = \frac{(x-a)^{p-1} \cdot (b-x)^{q-1}}{(b-a)^{p+q-1}}. \tag{14}$$

In [6] the parameters p and q of the given random variable are found, as well as the interval [a,b] in which it is defined. Therefore, we assume that the distribution density of the random variable (11) has an estimation that is accurate to the parameters.

Next, consider the case of n distinct paths. In general, let a random variable ξ be a function of random variables $\xi_1, \ldots, \xi_n$. We assume on the basis of [13] that each of these random variables has a beta distribution in the interval $[a_i, b_i]$. Then its numerical characteristics will be determined as follows [10]:

$$M(\xi) = \int\limits_{a_1}^{b_1} \ldots \int\limits_{a_n}^{b_n} \max(x_1, \ldots, x_n) f_\xi(x) dx_1 .. dx_n \tag{15}$$

and

$$D(\xi) = \int\limits_{a_1}^{b_1} \ldots \int\limits_{a_n}^{b_n} (\max(x_1, \ldots, x_n))^2 \cdot f_\xi(x_1, \ldots, x_n) dx_1 .. dx_n - (M(\xi))^2 \tag{16}$$

In comparison with formulas (7) and (8), the following difficulties arise in the n-dimensional case:

- the function of the module of several arguments becomes much more complicated (from the point of view of integration);
- it is obvious that the density function will become more cumbersome.

Next, we find the distribution density of the random variable ξ described by the formula (5). Since $\xi_1,\ldots,\xi_n$ are independent, the distribution function of ξ will have the following form:

$$F_\xi(x) = P(\xi < x) = P(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n) = P(\xi_1 < x_1)P(\xi_2 < x_2)\ldots P(\xi_n < x_n) =$$
$$= F_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2) \cdot \ldots \cdot F_{\xi_n}(x_n) \tag{17}$$

Taking the derivative, we obtain:

$$f_\eta(x) = f_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2) \cdot \ldots \cdot F_{\xi_n}(x_n) + f_{\xi_2}(x_1) \cdot F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n) + \ldots + f_{\xi_n}(x_n) \cdot F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_{n-1}}(x_{n-1}). \tag{18}$$

We substitute in the formula (17) the values of the function and density of beta-distribution. The first term is recorded:

$$f_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2) \cdot \ldots \cdot F_{\xi_n}(x_n) = \frac{(x_1 - a_1)^{p_1 - 1} \cdot (b_1 - x_1)^{q_1 - 1}}{(b_1 - a_1)^{p_1 + q_1 - 1}} \times \int_{a_2}^{x_2} (t - a_2)^{p_2 - 1} \cdot (b_2 - t)^{q_2 - 1} dt \times$$
$$\times \frac{1}{(b_2 - a_2)^{p_2 + q_2 - 1}} \cdot \ldots \times \frac{1}{(b_2 - a_2)^{p_n + q_n - 1}} \times \int_{a_n}^{x_n} (t - a_n)^{p_n - 1} \cdot (b_n - t)^{q_n - 1} dt \tag{19}$$

The rest of the terms will have the same form. As can be seen from the results, the analytical approach to finding the numerical characteristics of a random variable (5) is associated with a number of difficulties. They are called in particular:

- cumbersome expressions (19) (as components of integral expressions (15) and (16));
- difficulties arising when integrating the density (14);
- difficulties arising in the analytical description of the function of the maximum of n values.

In this regard, it can be concluded that to find the numerical characteristics of the formulas (15) and (16) it is necessary to connect a numerical apparatus.

5. Conclusions

The aim of the work was to describe a random variable characterizing the duration of servicing a complex of series-parallel operations and finding its numerical characteristics. The main results are listed below.

The existing approach for the estimation of numerical characteristics of service duration in stochastic systems is analysed by means of a computational experiment. Compared to estimates, the PERT method revealed that the actual expectation exceeds the estimated valuation, and the variance is lower than the planned evaluation. The results indicate the need to develop approaches for estimating the numerical characteristics of the service duration, characterized by high accuracy.

The analytical description of the investigated random variable based on the use of the probability theory apparatus is proposed. In particular, it was shown that it can be described by formula (7).

Analytical approaches to finding the numerical characteristics of the described random variable are investigated. In general, they are described by formulas (15) and (16). As a result, it is concluded that it is advisable to search for numerical approaches (or a combination of numerical and analytical approaches) to solve this problem.

References
[1]   Kofman A and Debazey G 1968 *Network planning methods* 517
[2]   Ma W, Che Y and Huang H 2016 Resource-constrained project shceduling problem with uncertain duration and renewable recources *International Journal of Machine Learning and*

   Cybernetics **7(4)** 613-21

[3] Artigues C, Demassey S and Neron E 2010 Resource-Constrained Project Shceduling: Models, Algorithms, Extensions and Applications 308

[4] Gupta A and Nadarajah S 2004 Handbook of Beta Distribution and Its Applications 579

[5] Golenko-Ginzburg D, Burkov V and Ben-Yair A 2011 Planning and controlling multilevel man-machine organization systems under random disturbances 448

[6] Oleynikova S 2013 Computer experiment for the analysis of the distribution law of the random quantity that determines the duration of the project in problems of network planning and management *Economics and Management of Control Systems* **9(3)** 90-6

[7] Gumbel E 1965 *Statistics of extrems* 452

[8] Nevzorov V 2000 *Records: mathematical theory* 256

[9] Haan L De and Ferreira A 2006 *Extreme value theory* 418

[10] Ventcel E 1999 *Theory of probabilities* 576

[11] Johnson N, Kotz S and Balakrishnan N 1995 *Continuous univariate distributions* **2** 752

[12] Whittaker E 2012 *The Calculus of Observations – A Treatise on Numerical Mathematics* 412

[13] Oleinikova S and Kravets O 2013 Approximation of the distribution law of the sum of random beta values *International Journal of Information Technologies and Security* **9(3)** 53-64