Adv Syst Sci Appl 2024; **02**:1–18
Published online at https://ijassa.ipu.ru.

Correlations of the Sojourn Times of Subtasks in Fork-Join Queueing Systems with $M|M|1$-type Subsystems

Anastasia V. Gorbunova[1*], Alexey V. Lebedev[2]

[1] *V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia*
[2] *Lomonosov Moscow State University, Moscow, Russia*

Abstract: The article analyzes the relationship between the residence times of subtasks in the fork-join subsystems of the queuing system. When arriving the fork-join system, the task is divided into subtasks, each of which is serviced in its own subsystem, the task is considered serviced after the completion of all subtasks that originally comprised it servicing. There is a dependence between the residence times of subtasks in subsystems, which affects the main performance indicators of the system, for example, the response times, which greatly complicates their analysis. The paper examines the characteristics of the existing dependence. In particular, with the help of generating functions and the Laplace-Stieltjes transformation, exact expressions for the Pearson and Spearman correlation coefficients are obtained. In addition, using a combination of several methods, including the Nelder-Mead optimization method, the estimation of the Kendall correlation coefficient was obtained, and the model for the response time estimation based on the resulting correlation coefficients was described. Despite the many works on the study of fork-join queuing systems, there are practically no articles devoted to the correlation analysis of the temporal metrics of the model. Therefore, this article can become one of the first in this area, laying the foundation for the study of the correlation dependence and its influence on the performance parameters of the model, not only for the classical case of $M|M|1$ subsystems, but also for more complex architectures of this system.

Keywords: fork-join queueing system, system with parallel service of tasks, correlation coefficient, Pearson correlation, Spearman correlation, Kendall correlation, generating function, meta-Gaussian model, simulation modeling.

1. INTRODUCTION

The article studies the classical fork-join queuing system (QS) with $M|M|1$-type subsystems. Upon arriving the system, the task is divided into several subtasks (the number of which is equal to the number of subsystems), each of which rises in turn for service. After servicing, the subtasks are again combined into a whole task, and the collected task leaves the system. Thus, the residence time of the entire task (the response time) is determined by the maximum of the residence times of its constituent subtasks.

The described mechanism for the functioning of the fork-join QS is suitable for modeling multitasking processes of various nature. Due to the fact that the division of a complex combined task into several subtasks is one of the ways to save time and resources, the analysis of fork-join QS is still an urgent task.

First of all, if we talk about the practical application of this model, it is worth mentioning the processes occurring in information and computing systems. For example, we can talk

*Corresponding author: avgorbunova@list.ru

about computing platforms that support data-intensive applications, and in general about various high-performance environments for big data analysis, which are based on parallel or distributed structures (MapReduce technology) [12,21,27,35]. There are also many examples of parallel structures in manufacturing systems, such as the process of assembling multi-item orders in warehouses of one or more suppliers [6, 19, 31]. In the field of healthcare, when patients are admitted or discharged, medical studies are carried out, which are distinguished by their duration [2, 15, 20]. In the banking sector, consideration of a loan application can be carried out by several departments simultaneously, which also leads to parallelization of this procedure [17].

In all systems described, the response time plays a key role. This characteristic is critical and quite often appears in agreements on the quality of services provided. Therefore, predicting both the average response time and the quantiles of this random variable allows one to make a qualitative prediction of one of the most important indicators of system performance.

The main reason for the complexity of the analysis of fork-join QS is the existing dependence between the times subtasks spend in subsystems. This dependence arises due to the commonality of the moment when subtasks appear in the system itself, since they are constituent elements of one task, which was divided into parts at the time of arriving into the system. The dependence between the sojourn times of subtasks is a distinctive feature of the fork-join QS with K subsystems $M|M|1$ (the analyzed type of QS in this particular case) from K parallel QS $M|M|1$. Therefore, it is of particular interest to estimate the correlation coefficients between the residence times of subtasks.

It is also important to estimate correlation coefficients because this expands the rather limited range of known methods for analyzing system response time. In particular, the article presents one of the variants of the model based on the known value of the Spearman correlation coefficient, which makes it possible to evaluate various response time characteristics by using the example of mathematical expectation based on the meta-Gaussian model. This model is not the only one possible in this case, but it serves as a clear example of the viability of the proposed approach for assessing the characteristics of a fork-join queueing system using correlation coefficients.

Despite the fact that the fork-join QS variant with $M|M|1$ subsystems is considered the easiest to study, accurate results were obtained only for the average response time in the case with two subsystems ($K = 2$) [23]. For $K > 2$, only approximations of the mean response time and its variance of varying degrees of accuracy [32–34] were obtained. There are also known publications in which estimates of the time moments of the task's stay in the QS, as well as estimates of high-level percentiles (for example, the 99th percentile) of the response time distribution in the case of more complex types of QS subsystems [4, 25, 26, 29] . By more complex types of subsystems we mean QS with non-Poisson input flows and non-exponential distribution of service time. Along with the classical analytical methods of queuing theory, fork-join systems have been studied empirically, using elements of the theory of order statistics and even using machine learning methods [7,9,36]. Various generalizations and modifications of fork-join systems [3, 28, 30] are also considered.

We also note that the authors previously studied the dependence of the maximum residual service times in two subsystems for a fork-join system with an infinite number of servers and general service times, for which the copulas and Blomquist coefficients [8] were found.

As for the evaluation of the correlation coefficients between the residence times of subtasks, there are practically no studies in this direction, and the authors could not find publications in which such an analysis would be carried out even in the case of subsystems of the $M|M|1$ type. In [10], the authors studied systems with Pareto service and obtained empirical formulas for the Pearson, Spearman and Kendall correlation coefficients (along with mean and standard deviation estimates) that give a good fit over a wide range of parameters. In the same work, it was possible to derive not estimates of the Pearson and

Spearman correlation coefficients, but exact formulas. An approximation for the Kendall correlation coefficient is also obtained.

The article provides the detailed description of the approach to the derivation of expressions for the correlation coefficients between the residence times of a pair of subtasks in the respective subsystems. By itself, the approach for deriving Pearson and Spearman correlations is classical and is based on the theory of generating functions and Laplace-Stieltjes transformations, while considering a system for which the number of subsystems is two ($K = 2$). To evaluate the Kendall correlation, the combined approach is used, which includes the Nelder-Mead optimization method and graphical analysis.

An interesting phenomenon has been discovered: at low load, the Pearson correlation coefficient is greater, and at high load, Spearman's. In this case, the Kendall correlation coefficient is always less than both of them.

It should be noted that the value of the correlation coefficients between any pair of subtasks is not affected by the number of subsystems, i. e. the value of the correlation coefficient does not depend on K and will be the same for any paired combination of subtasks from two different subsystems with fixed input flow and service parameters, and hence also true for all $K > 2$.

The article is organized as follows. Section 2 describes the mathematical model of fork-join QS. Sections 3–5 present a description of approaches to finding each of the three Pearson, Spearman and Kendall correlation coefficients, respectively, with elements of numerical analysis using a specific example of numerical data to confirm the obtained analytical result comparing with the results of simulation modeling in the Python software environment. In Section 6, the limiting two-dimensional distribution for the normalized residence times of subtasks under high load is obtained and all three correlation coefficients are compared. Section 7 describes the meta-Gaussian model for the average response time estimation using the Spearman correlation coefficient, also the numerical example is presented. The Conclusion (Section 8) summarizes some results and outlines ways for further research.

2. MATHEMATICAL MODEL OF FORK-JOIN QS

Consider a classical fork-join queuing system with $K \geqslant 2$ subsystems (Fig. 2.1). Tasks enter the system according to a Poisson process with a rate $\lambda > 0$. Upon arriving the system, a task is instantly divided into K subtasks, each of which is queued for service to one of K subsystems or immediately starts to be serviced if the queue is empty. All subsystems have one server for servicing subtasks and buffer of unlimited capacity. The service time on each server has an exponential distribution with the parameter $\mu > 0$. Thus, the subsystems are K

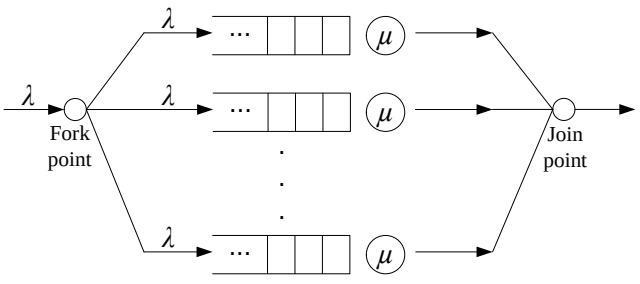

Fig. 2.1. Fork-join model of a queuing system with K subsystems of type $M_\lambda|M_\mu|1$.

identical QSs of type $M_\lambda|M_\mu|1$. A task is considered serviced only after all its parts, i. e. all its constituent subtasks, have been serviced.

Next, we turn to the consideration of the case when $K = 2$, since the total number of subsystems has no effect on the pairwise dependence of the times of sojourn in subsystems.

Consequently, the random residence time of a task in the QS (or the response time) of the R system is the maximum of two random residence times of the subtasks ξ_i, $i = 1.2$, in the corresponding subsystems

$$R = \max\{\xi_1, \xi_2\}.$$

The functioning of the system is described by the Markov process $X(t) = \{X_1(t), X_2(t)\}$, where $X_i(t)$ is the number of subtasks in the i-th subsystem, $i = 1, 2$. Then the set of process states can be written as follows $X = \{(i, j), i \geqslant 0, j \geqslant 0\}$. Further, let p_{ij} be the stationary probability that the first subsystem contains i customers and the second subsystem has j customers. The system ergodicity condition is standard and common for both queuing subsystems, i. e. $\rho = \lambda/\mu < 1$.

Now we introduce a generating function for the number of subtasks in the system

$$P(z, w) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} z^i w^j p_{ij}. \tag{2.1}$$

According to [5] it will take the form

$$P(z, w) = \frac{N(z, w)}{Q(z, w)},$$

where for $\lambda = 1$ (which, without loss of generality, we will assume everywhere below)

$$N(z, w) = \mu z(w - 1)P(z, 0) + \mu w(z - 1)P(0, w),$$

$$Q(z, w) = (1 + 2\mu)zw - \mu w - \mu z - z^2 w^2,$$

$$P(z, 0) = \frac{(\mu - 1)^{3/2}}{\mu(\mu - z)^{1/2}} = \frac{(1 - \rho)^{3/2}}{(1 - \rho z)^{1/2}}, \quad P(0, w) = \frac{(1 - \rho)^{3/2}}{(1 - \rho w)^{1/2}},$$

then

$$P(z, w) = \frac{z(w - 1)P(z, 0) + w(z - 1)P(0, w)}{(2 + \rho)zw - w - z - \rho z^2 w^2}. \tag{2.2}$$

In the following sections, we will describe the sequence of actions required to calculate the correlation coefficients of several types.

3. PEARSON CORRELATION COEFFICIENT

The dependence between the residence times of subtasks in the fork-join subsystems of a system with parallel servicing of tasks arises due to the common moments of tasks arrival in these subsystems. Up or down fluctuations of the input flow (according to the number of arrivals for some time) lead to an increase or decrease in the length of queues in subsystems and, accordingly, an increase or decrease in the residence time of subtasks of one task in subsystems.

Data visualization makes it possible to demonstrate the dependence of the sojourn times in subsystems. So, in fig. 3.2 shows the values of ξ_1 and ξ_2 obtained by simulation in the case of a fork-join QS and for the case of two parallel QSs $M|M|1$ with identical values of the parameters λ and μ. The number of pairs of points (ξ_1, ξ_2) is the same in both cases and is one million. Of course, such a number of points is not enough to obtain a good estimate of the correlation coefficient, but it is quite acceptable to illustrate the presence of a difference in the behavior of random variables for two options for the functioning of systems.

From the theory it is clear and clearly visible that for independent residence times (Fig. 2b) the joint density level lines have the form $x_1 + x_2 = const$ (since partial distributions

are exponential), while for dependent residence times (Fig. 2a) it can be seen that the lines have a bulge from the origin, which is the greater, the greater the load. This reflects a greater likelihood of jointly larger values than with independence.

Note that visually the nature of the dependence is not similar to the classical case in statistics, when there is a functional dependence of quantities (linear or monotonic), on which random noise is superimposed. Therefore, the question arises of how the known correlation coefficients will be able to capture (reflect) this dependence.

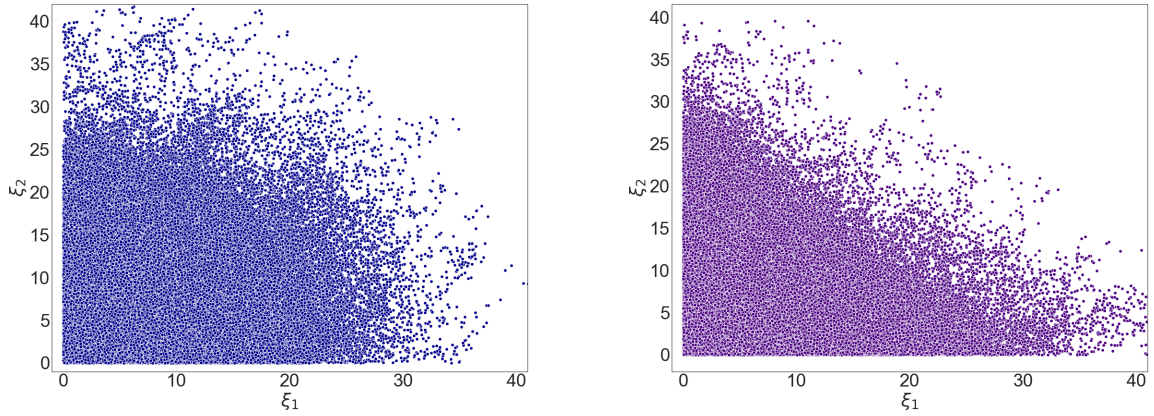

Fig. 3.2. Illustration of presence/absence of dependence between random variables ξ_1 and ξ_2 at $\rho = 0.8$ in case of a) fork-join QS with two subsystems $M|M|1$; b) two parallel functioning QS $M|M|1$.

To calculate the Pearson correlation coefficient, we will use the classical tools from queuing theory. Consider the Laplace-Stieltjes transformation (LST) for the residence times of subtasks in subsystems

$$\varphi(s,t) = \int_0^\infty \int_0^\infty e^{-sx} e^{-ty} v(x,y) dx dy, \tag{3.3}$$

where

$$v(x,y) = \frac{\partial^2 V(x,y)}{\partial x \partial y}$$

is the two-dimensional distribution density of the residence times of subtasks in subsystems, $V(x,y) = P(\xi_1 < x, \xi_2 < y)$.

Next, we carry out the standard arguments. If, when a task enters the system, it encounters i subtasks in the first QS and j subtasks in the second QS, then the waiting time in the queue of each of its two related elements in the corresponding subsystems will consist of the sum of the random service times of all preceding them in queue i or j of subtasks, respectively, as well as the service time of the newly arrived element itself. Taking into account the fact that the service time of one subtask has an exponential distribution with the parameter μ, then the sum consisting of $(i+1)$ and $(j+1)$ of such random variables will have the Erlang distribution with the distribution function $E_{i+1}(x)$ and $E_{j+1}(x)$. Then the two-dimensional distribution function of the residence time of subtasks in subsystems will take the form

$$V(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E_{i+1}(x) E_{j+1}(y) p_{ij}. \tag{3.4}$$

Using the properties of the LST, taking into account the expression (3.4), we obtain that

$$\varphi(s,t)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\varepsilon_{i+1}(s)\varepsilon_{j+1}(t)p_{ij}=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left(\frac{\mu}{\mu+s}\right)^{i+1}\left(\frac{\mu}{\mu+t}\right)^{j+1}p_{ij},$$

where $\varepsilon_{i+1}(s)$ and $\varepsilon_{j+1}(t)$ — LST for $E_{i+1}(x)$. Thus, the LST of the residence times of subtasks in subsystems (3.3) can be represented as

$$\varphi(s,t)=P\big(\frac{\mu}{\mu+s},\frac{\mu}{\mu+t}\big)\cdot\frac{\mu}{\mu+s}\cdot\frac{\mu}{\mu+t},\tag{3.5}$$

where $P(\cdot,\cdot)$ is the generating function from (2.1).

Taking into account $\lambda=1$, we have

$$\varphi(s,t)=P\big(\frac{1}{1+\rho s},\frac{1}{1+\rho t}\big)\cdot\frac{1}{1+\rho s}\frac{1}{1+\rho t}.\tag{3.6}$$

The Pearson correlation coefficient between the residence times of subtasks in subsystems is determined by the following expression

$$r_p=\frac{E[\xi_1\cdot\xi_2]-E[\xi_1]E[\xi_2]}{\sqrt{Var[\xi_1]Var[\xi_2]}}.\tag{3.7}$$

Taking into account the fact that, as is known, the residence time of an task in a system of type $M_\lambda|M_\mu|1$ has an exponential distribution with the parameter $(\mu-\lambda)$, i. e.

$$E[\xi_1]=E[\xi_2]=\frac{1}{\mu-\lambda},\quad Var[\xi_1]=Var[\xi_2]=\frac{1}{(\mu-\lambda)^2},$$

the expression (3.6) with $\lambda=1$ is converted to the form

$$r_p=(\mu-1)^2E[\xi_1\cdot\xi_2]-1=\mu^2(1-\rho)^2E[\xi_1\cdot\xi_2]-1.\tag{3.8}$$

Therefore, to determine the correlation coefficient, it is necessary to calculate $E[\xi_1\cdot\xi_2]$. This can be done with LST $\varphi(s,t)$

$$E[\xi_1\cdot\xi_2]=\int_0^\infty\int_0^\infty xy\cdot v(x,y)dxdy=$$

$$=\frac{\partial^2\varphi(s,t)}{\partial s\partial t}\bigg|_{s=0,t=0}=\frac{\partial^2\big[P\big(\frac{1}{1+\rho s},\frac{1}{1+\rho t}\big)\cdot\frac{1}{1+\rho s}\frac{1}{1+\rho t}\big]}{\partial s\partial t}\bigg|_{s=0,t=0},$$

while

$$P\big(\frac{1}{1+\rho s},\frac{1}{1+\rho t}\big)\cdot\frac{1}{1+\rho s}\frac{1}{1+\rho t}=$$

$$=\frac{(1-\rho)^{3/2}\left(t\frac{\sqrt{\rho s+1}}{\sqrt{\rho s-\rho+1}}+s\frac{\sqrt{\rho t+1}}{\sqrt{\rho t-\rho+1}}\right)}{\rho^2s^2t+\rho^2st^2-\rho^2st+\rho s^2+2\rho st-\rho s+\rho t^2-\rho t+s+t}.$$

If we introduce the following notation

$$F(s)=\frac{(1-\rho)^{3/2}\sqrt{1+\rho s}}{\sqrt{1+\rho s-\rho}},$$

$$G(s,t) = \rho^2 s^2 t + \rho^2 s t^2 - \rho^2 s t + \rho s^2 + 2\rho s t - \rho s + \rho t^2 - \rho t + s + t,$$

then the formula can be written in a more compact form

$$P\Big(\frac{1}{1+\rho s}, \frac{1}{1+\rho t}\Big) \cdot \frac{1}{1+\rho s}\frac{1}{1+\rho t} = \frac{tF(s) + sF(t)}{G(s,t)}.$$

Taking derivatives will lead to the following expression

$$\frac{\partial^2}{\partial s \partial t}\left[\frac{tF(s) + sF(t)}{G(s,t)}\right] = \frac{f(s,t)}{g(s,t)},$$

where

$$f(s,t) = G^2(s,t)[F'(s) + F'(t)] + G'_s(s,t)G'_t(s,t)[2tF(s) + 2sF(t)] -$$
$$-G'_s(s,t)G(s,t)[F(s) + sF'(t)] - G'_t(s,t)G(s,t)[F(t) + tF'(s)] -$$
$$-G''_{st}(s,t)G(s,t)[tF(s) + sF(t)],$$
$$g(s,t) = G^3(s,t).$$

Next, we need to calculate

$$\frac{\partial^2 \varphi(s,t)}{\partial s \partial t}\bigg|_{s=0,t=0} = \lim_{\substack{s\to 0\\ t\to 0}} \frac{\partial^2 \varphi(s,t)}{\partial s \partial t} = \lim_{\substack{s\to 0\\ t\to 0}} \frac{f(s,t)}{g(s,t)}.$$

Moreover, note that the functions $f(s,t)$ and $g(s,t)$ are infinitesimal of the third order as $(s,t) \to (0,0)$, i. e.

$$\lim_{\substack{s\to 0\\ t\to 0}} f(s,t) = \lim_{\substack{s\to 0\\ t\to 0}} g(s,t) = 0,$$
$$f'_s(0,0) = f'_t(0,0) = g'_s(0,0) = g'_t(0,0) = 0,$$
$$f''_{ss}(0,0) = f''_{tt}(0,0) = f''_{st}(0,0) = g''_{ss}(0,0) = g''_{tt}(0,0) = g''_{st}(0,0) = 0,$$
$$f^{(3)}_{s^{n_1}t^{n_2}}(0,0) \neq 0, \quad g^{(3)}_{s^{n_1}t^{n_2}}(0,0) \neq 0, \quad n_1 + n_2 = 3.$$

Therefore, for the existence of a double limit at the point $(0,0)$, it is necessary and sufficient that the following equality [13, 14] to be true:

$$\frac{f^{(3)}_{s^{n_1}t^{n_2}}(0,0)}{g^{(3)}_{s^{n_1}t^{n_2}}(0,0)} = m, \quad m \neq 0, m \neq \pm\infty, \quad n_1 + n_2 = 3. \tag{3.9}$$

and the double limit itself will be equal to m from (3.9)

$$\lim_{\substack{s\to 0\\ t\to 0}} \frac{f(s,t)}{g(s,t)} = m.$$

After appropriate calculations, we get that

$$f^{(3)}_{s^{n_1}t^{n_2}}(0,0) = \frac{3(1-\rho)\rho^2(4\rho - \rho^2 + 8)}{4}, \quad g^{(3)}_{s^{n_1}t^{n_2}}(0,0) = 6(1-\rho)^3, \quad n_1 + n_2 = 3.$$

Thus, we have

$$\frac{\partial^2 \varphi(s,t)}{\partial s \partial t}\bigg|_{s=0,t=0} = \lim_{\substack{s\to 0\\ t\to 0}} \frac{\partial^2 \varphi(s,t)}{\partial s \partial t} = \frac{f^{(3)}_{s^{n_1}t^{n_2}}(0,0)}{g^{(3)}_{s^{n_1}t^{n_2}}(0,0)} = \frac{\rho^2(4\rho - \rho^2 + 8)}{8(1-\rho)^2},$$

then

$$r_p = \mu^2(1-\rho)^2\frac{\rho^2(4\rho-\rho^2+8)}{8(1-\rho)^2} - 1 = \frac{\rho(4-\rho)}{8}. \tag{3.10}$$

The correctness of the (3.10) formula for the Pearson correlation coefficient between the residence times of any two subtasks in the corresponding subsystems is confirmed by a numerical experiment. With the help of simulation for the values $\lambda = 1$, $\rho \in [0.1, 0.9]$ with a step of 0.05, the values of r_p were calculated. The results of the comparison with the analytical expression (3.10) are shown in 3.3. Small deviations in the range of ρ values close to unity

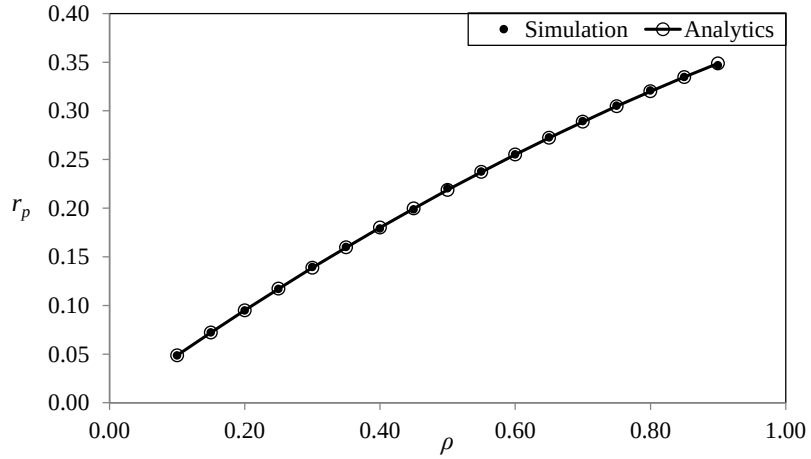

Fig. 3.3. Pearson correlation coefficient r_p

are explained as follows. An increase in the system load factor requires a significant increase in the duration of the run within one launch of the simulation model, since an increase in the load factor leads to an increase in the correlation of the data used to build a point estimate of the quantity under study. The features of the simulation fork-join QS and the construction of confidence intervals of the obtained estimates can be found in the work [11]. An increase in the duration of the run, in turn, leads to a significant increase in the time spent on the numerical experiment, which is not rational in this case, since the data match up to the third digit after the decimal point, and the module of the maximum relative error does not exceed 0.62%.

Based on the obtained results, we can conclude that the dependence between the residence times ξ_1 and ξ_2 in the subsystems is well reflected by the Pearson correlation coefficient, and it increases quadratically (with deceleration) with the load ρ.

4. SPEARMAN CORRELATION COEFFICIENT

Let's analyze the Spearman's correlation coefficient in order to get a complete picture of the relationship between the random residence times of subtasks ξ_1 and ξ_2 in subsystems.

In statistics, the Spearman correlation coefficient is calculated as the Pearson correlation coefficient applied to the ranks of observations (in ascending order) as a measure of the strength of a monotonic dependence (increasing or decreasing) between random variables, usually to test hypotheses about their independence.

From the point of view of probability theory, for continuous random variables, the Spearman correlation coefficient is conveniently defined as follows. Let random variables X_1 and X_2 have distribution functions F_1 and F_2, we consider $U_1 = F_1(X_1)$ and $U_2 = F_2(U_2)$, then the Spearman correlation coefficient X_1 and X_2 can be expressed as the Pearson correlation coefficient of U_1 and U_2 random variables [24, p. 170]: $r_s(X_1, X_2) = r_p(U_1, U_2)$.

The coefficient r_s, as well as r_p, takes values on the interval $[-1, 1]$.

To determine the coefficient r_s, let's make some preparatory calculations. Consider $\varphi(s,t)$ from (3.5), where s and t are multiplied by $(\mu-\lambda)$, then taking into account that $\rho = \lambda/\mu$, we get

$$\varphi((\mu-\lambda)s,(\mu-\lambda)t) = P(\frac{\mu}{\mu+(\mu-\lambda)s},\frac{\mu}{\mu+(\mu-\lambda)t})\cdot\frac{\mu}{\mu+(\mu-\lambda)s}\cdot\frac{\mu}{\mu+(\mu-\lambda)t} =$$

$$= P(\frac{1}{1+(1-\rho)s},\frac{1}{1+(1-\rho)t})\cdot\frac{1}{1+(1-\rho)s}\cdot\frac{1}{1+(1-\rho)t}.$$

Now let $\lambda = 1$, then we will substitute the corresponding expression for the generating function. We will get

$$\varphi((\mu-\lambda)s,(\mu-\lambda)t) =$$

$$= \frac{\frac{1}{1+(1-\rho)s}\left(\frac{1}{1+(1-\rho)t}-1\right)\cdot\frac{(1-\rho)^{3/2}}{(1-\frac{\rho}{1+(1-\rho)s})^{1/2}}+\frac{1}{1+(1-\rho)t}\left(\frac{1}{1+(1-\rho)s}-1\right)\cdot\frac{(1-\rho)^{3/2}}{(1-\frac{\rho}{1+(1-\rho)t})^{1/2}}}{(2+\rho)\frac{1}{1+(1-\rho)s}\cdot\frac{1}{1+(1-\rho)t}-\frac{1}{1+(1-\rho)s}-\frac{1}{1+(1-\rho)t}-\frac{\rho}{(1+(1-\rho)s)^2(1+(1-\rho)t)^2}}\cdot$$

$$\cdot\frac{1}{1+(1-\rho)s}\cdot\frac{1}{1+(1-\rho)t} = \frac{A(\rho)}{B(\rho)}\cdot\frac{1}{1+(1-\rho)s}\cdot\frac{1}{1+(1-\rho)t}.$$

Simplify the numerator $A(\rho)$

$$A(\rho) = \frac{1}{1+(1-\rho)s}\cdot\frac{1-1-(1-\rho)t}{1+(1-\rho)t}\cdot\frac{(1-\rho)^{3/2}}{\left(\frac{1+(1-\rho)s-\rho}{1+(1-\rho)s}\right)^{1/2}}+$$

$$+\frac{1}{1+(1-\rho)t}\cdot\frac{1-1-(1-\rho)s}{1+(1-\rho)s}\cdot\frac{(1-\rho)^{3/2}}{\left(\frac{1+(1-\rho)t-\rho}{1+(1-\rho)t}\right)^{1/2}} =$$

$$= \frac{-(1-\rho)^2}{(1+(1-\rho)s)(1+(1-\rho)t)}\cdot\frac{t(1+(1-\rho)s)^{1/2}}{(s+1)^{1/2}}+$$

$$+\frac{-(1-\rho)^2}{(1+(1-\rho)s)(1+(1-\rho)t)}\cdot\frac{s(1+(1-\rho)t)^{1/2}}{(t+1)^{1/2}} =$$

$$= \frac{-(1-\rho)^2\cdot C(\rho)}{(1+(1-\rho)s)(1+(1-\rho)t)}. \tag{4.11}$$

Simplify the denominator $B(\rho)$ with an additional factor

$$B(\rho)\cdot(1+(1-\rho)s)(1+(1-\rho)t) =$$

$$= 2+\rho-(1+(1-\rho)s)-(1+(1-\rho)t)-\frac{\rho}{(1+(1-\rho)s)(1+(1-\rho)t)} =$$

$$= \rho-(1-\rho)(s+t)-\frac{\rho}{(1+(1-\rho)s)(1+(1-\rho)t)} =$$

$$\frac{\rho(1-\rho)(s+t)+\rho(1-\rho)^2st-(1-\rho)(s+t)-(1-\rho)^2(s+t)^2-(1-\rho)^3(s+t)st}{(1+(1-\rho)s)(1+(1-\rho)t)} =$$

$$= \frac{-(1-\rho)^2\cdot\left((s+t)-\rho st+(s+t)^2+(1-\rho)(s+t)st\right)}{(1+(1-\rho)s)(1+(1-\rho)t)} =$$

$$= \frac{-(1-\rho)^2 \cdot D(\rho)}{(1+(1-\rho)s)(1+(1-\rho)t)} \tag{4.12}$$

Then

$$\varphi((\mu-\lambda)s,(\mu-\lambda)t) = \frac{C(\rho)}{D(\rho)},$$

where C and D are defined in (4.11) and (4.12), i. e.

$$\varphi((\mu-\lambda)s,(\mu-\lambda)t) = \frac{t(1+(1-\rho)s)^{1/2}(s+1)^{-1/2} + s(1+(1-\rho)t)^{1/2}(t+1)^{-1/2}}{(s+t)-\rho st+(s+t)^2+(1-\rho)(s+t)st}. \tag{4.13}$$

Recall that the sojourn times of subtasks in the fork-join subsystems of the QS have an exponential distribution with the parameter $\mu-\lambda$, i. e., $\xi_i \sim Exp(\mu-\lambda)$. Consider random variables $U_i = F(\xi_i)$, where $F_{\xi_i}(x) = 1-e^{-(\mu-\lambda)x}$, $x>0$, which will have uniform distribution on the interval $[0,1]$, i. e., $U_i \sim R[0,1]$, $i=1,2$. Then r_s — Spearman's correlation coefficient for random variables ξ_1 and ξ_2 will be Pearson's correlation coefficient for random variables U_1 and U_2 [24, p. 170], i. e.

$$r_s = \frac{E[U_1\cdot U_2]-E[U_1]E[U_2]}{\sqrt{Var[U_1]\cdot Var[U_2]}}.$$

Compute $E[U_1\cdot U_2]$:

$$E[U_1\cdot U_2] = E[(1-e^{-(\mu-\lambda)\xi_1})(1-e^{-(\mu-\lambda)\xi_2})] =$$

$$= 1 - E[e^{-(\mu-\lambda)\xi_1}] - E[e^{-(\mu-\lambda)\xi_2}] + E[e^{-(\mu-\lambda)\xi_1}e^{-(\mu-\lambda)\xi_2}]$$

Since $\eta_i = (\mu-\lambda)\xi_i \sim Exp(1)$, i. e., $p_{\eta_i}(x) = e^{-x}$, $x>0$, we have

$$E[e^{-(\mu-\lambda)\xi_i}] = E[e^{-\eta_i}] = \int\limits_0^\infty e^{-x}e^{-x}dx = \frac{1}{2}.$$

Then we get

$$E[U_1U_2] = 1-\frac{1}{2}-\frac{1}{2}+E[e^{-(\mu-\lambda)\xi_1}e^{-(\mu-\lambda)\xi_2}] = E[e^{-(\mu-\lambda)\xi_1}e^{-(\mu-\lambda)\xi_2}] =$$

$$= \int\limits_0^\infty\int\limits_0^\infty e^{-(\mu-\lambda)x}e^{-(\mu-\lambda)y}p_{\xi_1\xi_2}(x,y)dxdy =$$

$$= \int\limits_0^\infty\int\limits_0^\infty e^{-(\mu-\lambda)sx}e^{-(\mu-\lambda)sy}p_{\xi_1\xi_2}(x,y)\,dxdy\Big|_{s=1,t=1} =$$

$$= \varphi((\mu-\lambda)s,(\mu-\lambda)t)\Big|_{s=1,t=1}.$$

Now, substituting $s=1$ and $t=1$ into the formula (4.13), we obtain

$$E[U_1\cdot U_2] = \frac{\sqrt{2}\sqrt{2-\rho}}{8-3\rho}.$$

And given that $E[U_i] = 1/2$, and $Var[U_i] = 1/12$, $i = 1.2$, we can write the Spearman correlation coefficient in a final form

$$r_s = \frac{\frac{\sqrt{2}\sqrt{2-\rho}}{8-3\rho} - \frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{12}} = \frac{12\sqrt{2}\sqrt{2-\rho}}{8-3\rho} - 3. \tag{4.14}$$

Simulation modeling confirms the correctness of the (4.14) formula for the Spearman

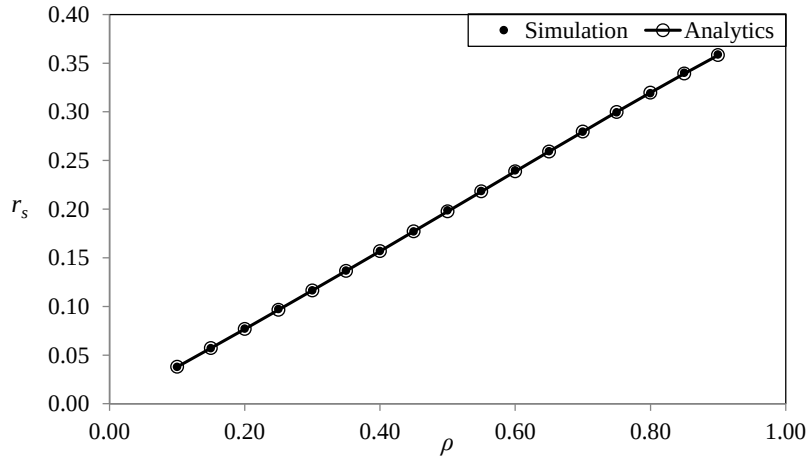

Fig. 4.4. Spearman correlation coefficient r_s

coefficient (Fig. 4.4).

Note that the Spearman correlation coefficient also increases with load in a non-linear manner, which however is visually closer to linear than for the Pearson correlation coefficient.

5. KENDALL CORRELATION COEFFICIENT

In this section, we construct an approximation for the Kendall correlation coefficient, since it has not yet been possible to derive an exact formula for it. The Kendall correlation coefficient, like the Spearman correlation coefficient, is a rank correlation coefficient. It also evaluates the nature of the monotonic dependence between random variables and the closeness of this connection [1, 18].

Statistically, in the sample, the proportion of pairs of observations of random vectors (from two random variables as components) is estimated, in which the components have the same order (monotonic character), i. e., for example, one component increases with the growth of the other, or vice versa, decreases with the growth of another. The Kendall correlation coefficient is estimated as the difference between the proportions of pairs of vectors for which the orders are the same and for which they differ. The formula for it can be written as follows

$$\hat{r}_k = 1 - \frac{4}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbf{1}\{[\xi_{1i} < \xi_{1j}] \neq [\xi_{2i} < \xi_{2j}]\}, \tag{5.15}$$

where $\mathbf{1}\{\cdot\}$ is the event indicator function $\{\cdot\}$, and (ξ_{1i}, ξ_{2i}), $1 \leq i \leq N$, is random sample of N vectors from random variables ξ_1 and ξ_2.

From the point of view of probability theory, the Kendall correlation coefficient of the random variables X and Y is defined as

$$r_k = \mathbf{E}\,\mathrm{sign}(X_1 - X_2)(Y_1 - Y_2),$$

where (X_1, Y_1), (X_2, Y_2) are independent random vectors distributed as (X, Y).

To approximate the Kendall correlation coefficient, we use a combination of several methods. First, we will carry out a graphical analysis of the data obtained using simulation modeling. After plotting the dependence of r_k on the system load ρ, you can see that, as in the case of r_s, this dependence is close to linear in appearance (compare Fig. 4.4 and Fig. 5.5), although in fact, as follows from the (4.14) formula, this is not the case. Assume for simplicity that r_k depends on ρ quadratically, i. e.

$$r_k \approx \rho(C_1 + C_2\rho).$$

Now, to find the unknown coefficients, we use the Nelder-Mead optimization method. In the process of optimization with respect to C, we will minimize the modulus of the relative approximation error

$$APE = \left| \frac{r_k - \widehat{r}_k}{r_k} \right| \cdot 100\%$$

between data generated by simulation (r_k — "true" values of Kendall's correlation coefficient) and predictions calculated by the proposed analytical formula ($\widehat{r}_k$ — Kendall's correlation coefficient estimate), with the initial values of the coefficients estimated from the graph.

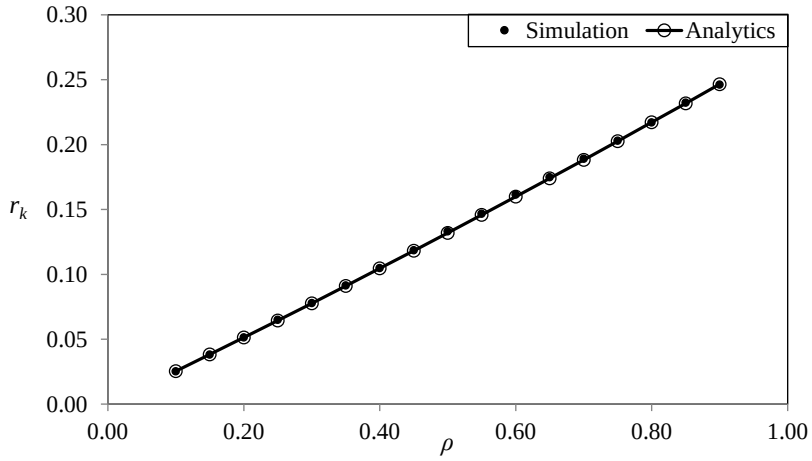

Fig. 5.5. Kendall correlation coefficient r_k

Table 5.1. Errors in approximating the Kendall correlation coefficient r_k by (5.16) for $\rho \in \{0.10, 0.15, ..., 0.90\}$

Estimated characteristic	Error types		
	Max APE, %	Min APE, %	MAPE, %
r_k	0.92828	0.09158	0.43138

As a result of optimization in the Python software environment, we obtain the values of the coefficients

$$C_1 \approx 0.25134, \quad C_2 \approx 0.02517.$$

Therefore, the final expression for r_k will look like

$$r_k \approx \rho(0.25134 + 0.02517\rho). \tag{5.16}$$

In the 5.5 figure and in the 5.1 table, you can compare the results of simulation modeling of the Kendall coefficient with the results of calculations using the analytical formula (5.16). As can be seen, the approximation error is not at all large and does not exceed 1

As for the values themselves, then, for example, $r_k \approx 0.2$ is true for $\rho = 0.8$, and this means that for about 60% of pairs of observations of the vectors (ξ_1, ξ_2) from the entire set of their the order (character of monotonicity) coincides, but for about 40% of the pairs (ξ_1, ξ_2) it does not. In this case, each vector corresponds to a task and consists of the sojourn times of its subtasks.

Thus, the nature of the dependence of the Spearman and Kendall correlation coefficients on the load in both cases is close to linear, but the Kendall coefficient takes smaller values.

6. LIMITING TWO-DIMENSIONAL DISTRIBUTION AND RATIOS OF CORRELATION COEFFICIENTS

From the obtained results, it can be seen that all correlation coefficients have some limits at high load ($\rho \to 1$), and these limits are different from both 0 and 1, which suggests that they have meaningful meaning.

Note that from the formula (4.13) one can obtain the limit

$$\lim_{\rho \to 1} \varphi((\mu - \lambda)s, (\mu - \lambda)t) = \frac{t(s+1)^{-1/2} + s(t+1)^{-1/2}}{s + t - st + (s+t)^2}, \tag{6.17}$$

and this is the LST of some two-dimensional distribution. Namely, this is the limit distribution of normalized sojourn times

$$\eta_1 = (\mu - \lambda)\xi_1, \quad \eta_2 = (\mu - \lambda)\xi_2,$$

whose joint distribution is described by the LST $\varphi((\mu - \lambda)s, (\mu - \lambda)t)$, with the random variables η_1 and η_2 individually are always equally distributed (they are standard exponentials), and the dependence between them is determined by the load of ρ.

All limit values of the correlation coefficients are thus the values of the correlation coefficients for the limit distribution (6.17). Let's write them out explicitly[†]:

$$r_p = \frac{3}{8} = 0.375, \quad r_s = \frac{12\sqrt{2}}{5} - 3 \approx 0.394, \quad r_k \approx 0.276.$$

In the figure 6.6 we also present a joint graph of all three coefficients (with the addition of limit values). As can be seen from the figure, the lines of the Pearson and Spearman correlation coefficients intersect at one point (except zero). The coordinates of the intersection point can be found by equating the expressions (3.10) and (4.14). After simplification, we obtain the equation

$$9\rho^6 - 120\rho^5 + 160\rho^4 + 2752\rho^3 - 6080\rho^2 + 3072\rho = 0.$$

The numerical solution of this equation allows us to find the only root that belongs to the interval $\rho \in (0, 1)$, namely $\rho \approx 0.803146$. Thus, the intersection point has the following approximate coordinates: $(0.803146, 0.320943)$. To the left of it is more Pearson's correlation coefficient, and to the right — Spearman's.

7. META-GAUSSIAN MODEL

The calculation of the residence time correlation coefficients is not only of academic interest, but can also be useful for assessing system performance. Indeed, when it comes to selecting an

[†]The value of r_k is an estimate according to the formula (5.16).

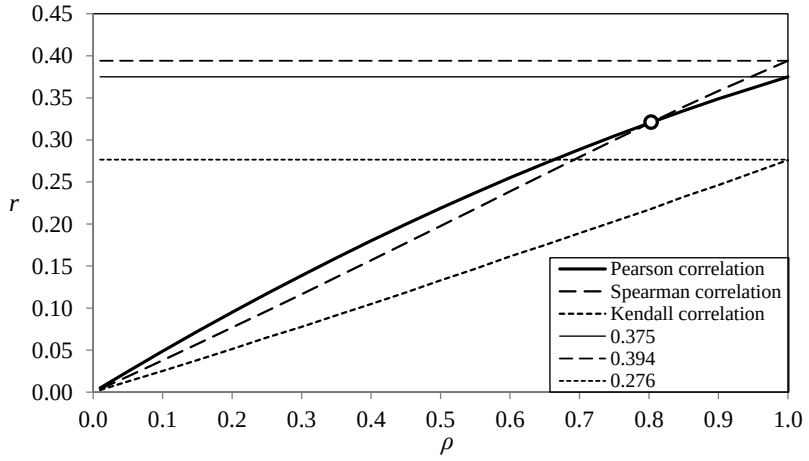

Fig. 6.6. Coefficients of correlation and their limit values

approximate model of the dependence of residence times, such a model can be parameterized by one of the correlation coefficients.

Next, we will consider the meta-Gaussian model based on reducing an arbitrary distribution to a multivariate normal one. Such models are used in financial mathematics [22, ch. 5, 9], hydrology [16], etc. At the same time, we do not claim that this model is optimal in this case, but we present it only as an example and plan to pass on to more accurate models in the future.

Let the number of subsystems be $K \geq 2$, and the random variables ξ_i, $1 \leq i \leq K$, as before, be the residence times of subtasks from one task. All of them have an exponential distribution with parameter $(\mu - \lambda)$. Let's put

$$\zeta_i = \Phi^{-1}\left(1 - e^{-(\mu-\lambda)\xi_i}\right), \quad 1 \leq i \leq K,$$

where Φ^{-1} is the inverse function of the standard normal distribution. Then the random variables ζ_i, $1 \leq i \leq K$, have the standard normal distribution and

$$\xi_i = -\frac{1}{\mu - \lambda} \ln\left(1 - \Phi(\zeta_i)\right), \quad 1 \leq i \leq K.$$

Suppose that ζ_i, $1 \leq i \leq K$, have a joint multivariate normal distribution. Then, due to the symmetry of the system, any pair of quantities ζ_i and ζ_j, $i \neq j$, has the same Pearson correlation coefficient, which we denote as r. In addition, this pair has the same Spearman correlation coefficient r_s as the original pair ξ_i and ξ_j, since the Spearman correlation coefficient is preserved under continuous monotonically increasing transformations of random variables. For a multivariate normal distribution [22, Theorem 5.36, p. 215] it is true, that:

$$r_s = \frac{6}{\pi} \arcsin\frac{r}{2},$$

where we can find

$$r = 2\sin\frac{\pi r_s}{6}, \tag{7.18}$$

with known r_s, calculated by using the formula (4.14).

A set of random variables ζ_i with the required distributions and correlations can be obtained using the formulas:

$$\zeta_i = \sqrt{r}\varepsilon_0 + \sqrt{1-r}\varepsilon_i, \quad 1 \leq i \leq K,$$

where ε_i, $0 \le i \le K$, are independent standard normal random variables.

So for the response time

$$R_K = \max\{\xi_1, \ldots, \xi_K\}$$

we get the estimate

$$\widehat{R}_K = -\frac{1}{\mu - \lambda} \ln\left(1 - \Phi(\sqrt{r}\varepsilon_0 + \sqrt{1-r}\max\{\varepsilon_1, \ldots, \varepsilon_K\})\right). \tag{7.19}$$

Using formulas (4.14), (7.18) and (7.19), a simulation was carried out for the values $\lambda = 1$, $\rho \in [0.1, 0.9]$ with steps of 0.05 and K from 3 to 20, in order to estimate the average response time in comparison with the results of the simulation of a fork-join queuing system carried out by the authors earlier.

The errors of the obtained approximations are presented in the table 7.2.

Table 7.2. Errors in approximations of the average response time using the formula (7.19) (meta-Gaussian model) for values $K = 3, ..., 20$ and $\rho \in \{0.10, 0.15, ..., 0.90\}$

Evaluated characteristic	Types of errors		
	Max APE, %	Min APE, %	MAPE, %
R_K	4.18222	0.29829	2.46825

Note that $MaxAPE$ turns out to be close to the error of the classical Nelson-Tantavi formula [23] on a given set of parameter values (about 4%).

The model can be easily used to estimate not only the average response time, but also variance, quantiles, etc.

The disadvantages of the meta-Gaussian model include the fact that it requires simulation rather than provides an explicit formula, but this simulation is much simpler and faster than the simulation of the original fork-join queuing system.

Of course, more advanced models, which parameterized by correlation coefficients, are possible. If the model uses parameterization for other reasons, then it seems desirable that its predictions of the values of the correlation coefficients do not diverge too much from the actual ones.

8. CONCLUSION

The article presents exact analytical expressions for the Pearson and Spearman correlation coefficients between the residence times of subtasks in the fork-join subsystems of QS. These formulas were obtained using the classical method of generating functions and Laplace-Stieltjes transformations. The authors could not find sources where such studies were carried out, perhaps one of the reasons is the cumbersomeness of the calculations necessary to derive these formulas.

An approximate expression was obtained for the Kendall correlation coefficient, the approximation accuracy of which is quite high. To derive the Kendall correlation estimate, graphical analysis and the Nelder-Mead optimization method were used. The results obtained in the framework of a numerical experiment were compared with the data of simulation modeling, which confirmed their correctness.

It is shown that all coefficients increase with increasing load, with Pearson's correlation coefficient quadratic (with slowdown), and Spearman and Kendall's in a more complex non-linear manner (but close to linear). At a high load, they approach some limiting values due to the properties of the limiting two-dimensional distribution of the normalized residence times of subtasks. The coefficients were compared according to how well they capture the dependence (they take larger values). It has been established that the Kendall coefficient is

the worst, and as for the rest, there is a critical load value, below which the Pearson coefficient is better, and above it is the Spearman coefficient.

Formulas for correlation coefficients make it possible to meaningfully describe the existing relationship between the random variables of the residence times of subtasks, which in turn may allow further more accurate analysis of the response time characteristics of the entire system for the case when the number of subsystems is more than two, since for such systems only approximations of the average response time and its variance have been obtained with varying degrees of accuracy. As a simple example along this path, the paper presents the meta-Gaussian model. To describe the dependence and evaluate the characteristics, it is also planned to use the modern theory of copulas, with the fitting of copulas to the available data.

REFERENCES

1. Agresti, A. (2010). *Analysis of Ordinal Categorical Data (Second ed.)*. New York: John Wiley & Sons.
2. Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y. & Yom-Tov, Y. (2015). Patient flow in hospitals: a data-based queueing-science perspective, *Stochastic Systems*, **5**, 146–194. https://doi.org/10.1287/14-SSY153
3. Bushkova, T., Moiseeva, S., Moiseev, A., Sztrik, J., Lisovskaya, E. & Pankratova, E. (2022). Using Infinite-server Resource Queue with Splitting of Requests for Modeling Two-channel Data Transmission, *Methodology and Computing in Applied Probability*, **24**, 1753–1772. https://doi.org/10.1007/s11009-021-09890-6
4. Fiorini, P. M. (2015) Analytic approximations of fork-join queues, in: *IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Warsaw, Poland*, 966—971. https://doi.org/10.1109/IDAACS.2015.7341448.
5. Flatto, L. & Hahn, S. (1984). Two Parallel Queues Created by Arrivals with Two Demands I, *SIAM Journal on Applied Mathematics*, **44**, 1041–1053. https://www.jstor.org/stable/2101136
6. Gallien, J. & Wein, L. M. (2001). A Simple and Effective Component Procurement Policy for Stochastic Assembly Systems. *Queueing Systems*, **38**, 221–248. https://doi.org/10.1023/A:1010914600116
7. Gorbunova, A. V. & Vishnevsky, V. M. (2020). Estimating the response time of a cloud computing system with the help of neural networks, *Advances in Systems Science and Applications*, **20**, 105–112. https://doi.org/10.25728/assa.2020.20.3.926
8. Gorbunova, A. V. & Lebedev, A. V. (2020). Bivariate Distributions of Maximum Remaining Service Times in Fork-Join Infinite-Server Queues. *Problems of Information Transmission*, **56**, 73–90. https://doi.org/10.1134/S003294602001007X
9. Gorbunova, A. V. & Lebedev, A. V. (2022). Response Time Estimate for a Fork-Join System with Pareto Distributed Service Time as a Model of a Cloud Computing System Using Neural Networks, *Communications in Computer and Information Science*, **1552**, 318–332. https://doi.org/10.1007/978-3-030-97110-6_25
10. Gorbunova, A. V. & Lebedev, A. V. (2023). Nonlinear Approximation of Characteristics of a Fork-Join Queueing System with Pareto Service as a Model of Parallel Structure of Data Processing, *Mathematics and Computers in Simulation*, **214**, 409–428. https://doi.org/10.1016/j.matcom.2023.07.029
11. Gorbunova, A. V. & Lebedev, A. V. (2023). On Estimating the Characteristics of a Fork-Join Queueing System with with Poisson Input and Exponential Service Times, *Advances in Systems Science and Applications*, **23**, 99–114. https://doi.org/10.25728/assa.2023.23.2.1351
12. Hashem, I. A. T., Anuar, N. B., Gani, A., Yaqoob, I., Xia, F. & Khan, S. U. (2016). MapReduce: Review and open challenges, *Scientometrics*, **109**, 1–34. https://doi.org/10.1007/s11192-016-1945-y

13. Ivlev, V. V. (2002). Indeterminacies of Functions of Many Variables (Part I), *Mathematical Education*, **4**, 90–100.
14. Ivlev, V. V. (2003). Indeterminacies of Functions of Many Variables (Part II). *Mathematical Education*, **3**, 77–85.
15. Jiang, L. & Giachetti, R. E. (2008). A queueing network model to analyze the impact of parallelization of care on patient cycle time, Health Care Management Science, **11**, 248–261. https://doi.org/10.1007/s10729-007-9040-9
16. Kelly, K. S. & Krzysztofowicz, R. (1997). A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrol Hydraul*, **11**, 17–31. https://doi.org/10.1007/BF02428423
17. Kemper, B. & Mandjes, M. (2012). Mean sojourn times in two-queue fork-join systems: bounds and approximations. *OR Spectrum*, **34**, 723–742 https://doi.org/10.1007/s00291-010-0235-y
18. Kendall, M. (1938). A New Measure of Rank Correlation, *Biometrika*, **30**, 81–89. https://doi.org/10.2307/2332226
19. Ko, S. S. & Serfozo, R. F. (2004). Response times in M/M/s fork-join networks, *Advances in Applied Probability*, **36**, 854–871.
20. Lee, S.-Y., Chinnam, R. B., Dalkiran, E., Krupp, S. & Nauss, M. (2021). Proactive coordination of inpatient bed management to reduce emergency department patient boarding, *International Journal of Production Economics*, **231**, Article No. 107842 https://doi.org/10.1016/j.ijpe.2020.107842
21. Malekimajd, M., Ardagna, D., Ciavotta, M., Gianniti, E., Passacantando, M. & Rizzi, A. M. (2018). An optimization framework for the capacity allocation and admission control of MapReduce jobs in cloud systems, *The Journal of Supercomputing*, **74**, 5314–5348. https://doi.org/10.1007/s11227-018-2426-2
22. McNeil, A. J., Frey, R. & Embrechts, P. (2005). *Quantitative risk management*. Princeton, NJ: Princeton University Press.
23. Nelson, R. & Tantawi, A. N. (1988). Approximate analysis of fork/join synchronization in parallel queues Computers, *IEEE Transactions on Computers*, **37**, 739–743. https://doi.org/10.1109/12.2213
24. Nelsen, R. (2006). *An introduction to copulas*. Berlin, Germany: Springer.
25. Nguyen, M., Alesawi, S., Li, N., Che, H. & Jiang H. (2018). ForkTail: A black-box fork-join tail latency prediction model for user-facing datacenter workloads, in: *Proc. 27th Int. Symp. High-Perform. Parallel Distrib. Comput.*, 206–217.
26. Nguyen, M., Alesawi, S., Li, N., Che, H. & Jiang H. (2020). A Black-Box Fork-Join Latency Prediction Model for Data-Intensive Applications, *IEEE Transactions on Parallel and Distributed Systems*, **31**, 1983–2000. https://doi.org/10.1109/TPDS.2020.2982137
27. Oliveira, D. C. M., Liu, J. & Pacitti, E. (2019). *Data-Intensive Workflow Management: For Clouds and Data-Intensive and Scalable Computing Environments*. Morgan & Claypool.
28. Osipov, O. & Rogachko, E. (2020). Analysis of Centralized Splitting Fork-Join Queueing Systems with Heterogeneous Servers and Threshold Control Policy, in: *Proceedings of the Second International Workshop on Stochastic Modeling and Applied Research of Technology (SMARTY 2020)*, 76–88. https://ceur-ws.org/Vol-2792/paper5.pdf
29. Qiu, Z., Pérez, J. F. & Harrison P. G. (2015). Beyond the mean in fork-join queues: Efficient approximation for response-time tails, *Performance Evaluation*, **91**, 99–116. https://doi.org/10.1016/j.peva.2015.06.007
30. Rumyantsev, A. S., Dolgaleva, D. S. & Golovin, A. S. (2023). Steady-state performance analysis of multiserver queueing models with redundancy. *Program Systems: Theory and Applications*, **14**, 76–94. https://doi.org/10.25209/2079-3316-2023-14-1-55-94
31. Schol, D., Vlasiou, M. & Zwart, B. (2022). Large Fork-Join Queues with Nearly Deterministic Arrival and Service Times, *Mathematics of Operations Research*, **47**,

1335–1364. https://doi.org/10.1287/moor.2021.1171
32. Thomasian, A. (2014). Analysis of fork/join and related queueing systems, *ACM Computing Surveys*, **47**(2), 17:1–17:71. https://doi.org/10.1145/2628913
33. Varki, E., Merchant, A. & Chen, H. (2008). The $M|M|1$ fork-join queue with variable subtasks, unpublished, available online. https://www.cs.unh.edu/ṽarki/publication/2002-nov-open.pdf
34. Varma S. & Makowski, A. M. (1994). Interpolation approximations for symmetric fork-join queues, *Performance Evaluation*, **20**, 245–265. https://doi.org/10.1016/0166-5316(94)90016-7
35. Vianna, E., Comarela, G., Pontes, T., Almeida, J., Almeida, V., Wilkinson, K., Kuno, H. & Dayal, U. (2013). Analytical performance models for MapReduce workloads, *International Journal of Parallel Programming*, **41**, 495–525. https://doi.org/10.1007/s10766-012-0227-4
36. Vishnevsky, V. M. & Gorbunova, A. V. (2022). Application of machine learning methods to solving problems of queuing theory, *Communications in Computer and Information Science*, **1605**, 304–316. https://doi.org/10.1007/978-3-031-09331-9_24

