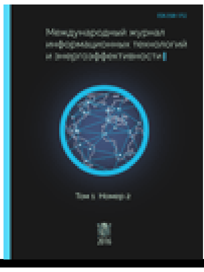
Подгорнов М.Д. Модель системы массового обслуживания «ТОЧНО-В-СРОК» С относительным приоритетом в обслуживании // Международный журнал информационных технологий и энергоэффективности. – 2025. – Т. 10 № 4(54) с. 135–141

ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и энергоэффективности

Сайт журнала: http://www.openaccessscience.ru/index.php/ijcse/



УДК 004.942

МОДЕЛЬ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ «ТОЧНО-В-СРОК» С ОТНОСИТЕЛЬНЫМ ПРИОРИТЕТОМ В ОБСЛУЖИВАНИИ

Подгорнов М.Д.
ФГБОУ ВО "УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ", Ульяновск, Россия, (432017, Ульяновская область, город Ульяновск, ул. Льва Толстого, д. 42), e-mail: maksimka_7373@mail.ru

В работе развивается семимартингальный (траекторный) подход к математическому описанию и моделированию систем массового обслуживания (СМО) «точно-в-срок» с приоритетами в обслуживании. Рассмотрена модель одноканальной СМО с относительным приоритетом. Показан переход от математической модели к итерационным формулам, по которым проводится имитационное моделирование.

Ключевые слова: Система массового обслуживания, семимартингальное описание, точно-в-срок, приоритет, точечный процесс, компенсатор, имитационное моделирование.

THE JUST-IN-TIME QUEUING SYSTEM MODEL WITH RELATIVE PRIORITY

Podgornov M.D.
ULYANOVSK STATE UNIVERSITY, Ulyanovsk, Russia, (432017, Ulyanovsk region, Ulyanovsk city, Lva Tolstoy str., 42), e-mail: maksimka_7373@mail.ru

The paper develops a semi-martingale (trajectory) approach to the mathematical description and modeling of just-in-time queuing systems (QS) with service priorities. The model of singlechannel QS with relative priority is considered. The transition from a mathematical model to iterative formulas, which are used for simulation, is shown.

Keywords: Queuing System, semi-martingale description, just-in-time, priority, point process, compensator, simulation modeling.

Введение

В данной работе рассматривается достаточно новая для теории массового обслуживания система «точно-в-срок» с относительным приоритетом в обслуживании.

Алгоритм обслуживания "точно-в-срок" достаточно хорошо известен и применяется во различных областях. Эта модель фокусируется на высоком уровне сервисного обслуживания, минимизации времени ожидания и эффективном распределении ресурсов, что позволяет повышать общую производительность и удовлетворенность клиентов (см., к примеру, работы [1-2]).

Несмотря на широкое применение данной концепции в различных областях, модели систем массового обслуживания «точно-в-срок» находятся на довольно низком уровне развития. Это касается как имитационных, так и математических моделей. Однако, применение таких моделей необходимо при решении задач оптимального управления, так как они позволяют принимать оптимальные решения на основе анализа вероятностных характеристик системы, учитывая различные факторы неопределенности. Цель исследования заключается в разработке стохастического описания СМО «точно-в-срок» с относительным приоритетом (приоритетная заявка ожидает окончания обслуживания текущей заявки и после этого встает на внеочередное обслуживание), которое было бы подходящим как для аналитических методов, так и для компьютерного моделирования.

Для математического описания СМО использован аппарат точечных (считающих) процессов и их компенсаторов. С данным траекторным подходом при описании СМО можно ознакомиться, например, по работам [3-5]. Для контроля приоритета использован подход, основанный на регулировании размеров очередей.

Постановка задачи

Рассмотрим одноканальную СМО, в которую поступают заявки двух типов: первый тип – часто поступающие, но менее важные заявки, второй тип – более важные, но поступающие реже. Заявки обоих типов поступают независимо друг от друга и образуют простейшие потоки с интенсивностями $\lambda_1 > 0$ и $\lambda_2 > 0$. С момента начала обслуживания заявки, оператор должен завершить ее обработку за определенный отрезок времени, определяемый параметром $\tau_1 > 0$ для заявок первого типа и параметром $\tau_2 > 0$ для заявок второго типа. Так как заявки второго типа являются более важными, они имеют относительный приоритет в обслуживании, то есть встают на обслуживание после обработки текущей заявки не зависимо от количества заявок первого типа в системе. Общая очередь в данном случае, очевидно, не эффективна, поэтому для заявок каждого типа организована отдельная очередь. (Рисунок 1).

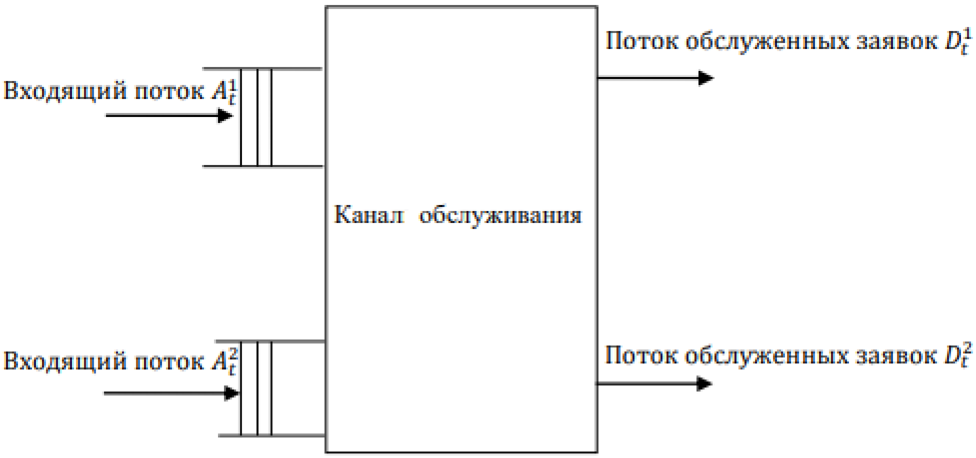


Рисунок 1 - Схема СМО

Математическая модель

Для описания работы систем введем считающие процессы A^1, A^2, D^1, D^2 где $A^1 = (A_t^1)_{t\geq 0}$ – число заявок первого типа, поступивших в СМО за время $t \geq 0$, $A_0^1 = 0$, $A^2 = (A_t^2)_{t\geq 0}$ – число заявок второго типа, поступивших в СМО за время $t \geq 0$, $A_0^2 = 0$, $D^1 = (D_t^1)_{t\geq 0}$ – число обслуженных заявок первого типа за время $t \geq 0$, $D_0^1 = 0$ и $D^2 = (D_t^2)_{t\geq 0}$ – число обслуженных заявок второго типа за время $t \geq 0$, $D_0^2 = 0$ Точечные процессы A^1, A^2, D^1 и D^2 определяются своими компенсаторами $\widetilde{A^1} = (\widetilde{A_t^1})_{t\geq 0}$, $\widetilde{A^2} =$

$(\widetilde{A_t^2})_{t\geq 0}, \widetilde{D^1} = (\widetilde{D_t^1})_{t\geq 0}$ и $\widetilde{D^2} = (\widetilde{D_t^2})_{t\geq 0}$ в соответствии с разложением Дуба-Мейера для субмартингалов [4]:

$$A_t^1 = \widetilde{A_t^1} + m_t^{A^1}, \tag{1}$$

$$A_t^2 = \widetilde{A_t^2} + m_t^{A^2}, \tag{2}$$

$$D_t^1 = \widetilde{D_t^1} + m_t^{D^1}, \tag{3}$$

$$D_t^2 = \widetilde{D_t^2} + m_t^{D^2}, \tag{4}$$

где $\widetilde{A^1}, \widetilde{A^2}, \widetilde{D^1}$ и $\widetilde{D^2} -$ неубывающие предсказуемые процессы, $m_t^{A^1}$, $m_t^{A^2}, m_t^{D^1}$ и $m_2^D -$ мартингалы.

Для рассматриваемой системы, компенсаторы процессов $A^1 = (A_t^1)_{t\geq 0}$ и $A^2 = (A_t^2)_{t\geq 0}$ определяются следующими соотношениями:

$$\widetilde{A_t^1} = \lambda_1 t,\ \lambda_1 > 0, \tag{5}$$

$$\widetilde{A_t^2} = \lambda_2 t,\ \lambda_2 > 0, \tag{6}$$

где $\lambda_1, \lambda_2 > 0 -$ интенсивность поступления заявок первого и второго типов соответственно.

Компенсаторы для процессов $D^1 = (D_t^1)_{t\geq 0}$ и $D^2 = (D_t^2)_{t\geq 0}$ будут иметь вид:

$$\widetilde{D_t^1} = \int_0^t \mu_s^1\ ds\ , \tag{7}$$

$$\widetilde{D_t^2} = \int_0^t \mu_s^2\ ds\ , \tag{8}$$

где μ_t^1 и $\mu_t^2 -$ интенсивности обслуживания заявок двух типов. Определять их будем следующими соотношениями:

$$\mu_t^1 = \frac{1}{t_t^{o1} - t} \cdot I(t_t^{o1} > 0). \tag{9}$$

$$\mu_t^2 = \frac{1}{t_t^{o2} - t} \cdot I(t_t^{o2} > 0). \tag{10}$$

Здесь $I(\cdot) -$ индикаторная функция, $t_t^{o1} -$ время, к которому оператор стремиться завершить обработку заявки первого типа. Аналогично, $t_t^{o2} -$ время, к которому оператор стремиться закончить обработку заявки второго типа. Отметим, что в любой момент времени $t \geq 0, \mu_t^1 \geq 0,\ \mu_t^2 \geq 0.$

Опишем управление относительным приоритетом через регулирование значений параметров t_t^{o1} и t_t^{o2} и размеров очередей. Для заявок первого типа уравнения будут иметь вид:

$$dt_t^{o1} = (t + \tau_1) \cdot I(A_t^1 + A_t^2 - D_t^1 - D_t^2 = 0) dA_t^1 + (t + \tau_1 - t_t^{o1}) \cdot I(q_t^1 > 0, q_t^2 = 0) dD_t^1 + \\ +(t + \tau_1) \cdot I(q_t^1 > 0, q_t^2 = 0) dD_t^2 - t_t^{o1} \cdot I(q_t^1 = 0) dD_t^1 - t_t^{o1} \cdot I(q_t^2 > 0) dD_t^1, \tag{11}$$

где $q_t^1 -$ количество заявок первого типа в очереди в момент времени $t \geq 0$, $q_0^1 = 0, q_t^2 -$ количество заявок второго типа в очереди в момент времени $t \geq 0$, $q_0^2 = 0$. Для параметра q_t^1 можно написать следующее балансовое уравнение:

$$dq_t^1 = I(A_t^1 + A_t^2 - D_t^1 - D_t^2 > 0) dA_t^1 - I(q_t^1 > 0, q_t^2 = 0) dD_t^1 - I(q_t^1 > 0, q_t^2 = 0) dD_t^2, \tag{12}$$

т.е. очередь будет увеличивается на единицу, если в момент прихода новой заявки первого типа $(dA_t^1 = 1)$ оператор занят, и уменьшаться на единицу, если в момент окончания обслуживания текущей заявки ($dD_t^1 = 1$ или $dD_t^2 = 1$) в очереди нет заявок второго типа ($q_t^2 = 0$), а очередь заявок первого типа не пуста($q_t^1 > 0$).

Логика построения уравнения (11) такова. Во-первых, параметр t_t^{o1} принимает значение равное сумме текущего значения времени и параметра τ_1, если в момент прихода новой заявки первого типа ($dA_t^1 = 1$) оператор свободен, либо если в момент окончания обслуживания текущей заявки ($dD_t^1 = 1$ или $dD_t^2 = 1$) в очереди есть заявки первого типа ($q_t^1 > 0$) и отсутствуют заявки второго типа $q_t^2 = 0$. Во-вторых, обнуляется, если в момент окончания обслуживания заявки первого типа ($dD_t^1 = 1$) либо очередь заявок первого типа пуста ($q_t^1 = 0$), либо в очереди есть заявки второго типа ($q_t^2 > 0$).

Уравнения для заявок второго типа будут таковы:

$$dt_t^{o2} = (t+\tau_2)\cdot I(A_t^1 + A_t^2 - D_t^1 - D_t^2 = 0)dA_t^2 + (t+\tau_2)\cdot I(q_t^2 > 0)dD_t^1 + \\ +(t+\tau_2 - t_t^{o2})\cdot I(q_t^2 > 0)dD_t^2 - t_t^{o2}\cdot I(q_t^2 = 0)dD_t^2 \quad (13)$$

Балансовое уравнение для параметра q_t^2 будет следующим:

$$dq_t^2 = I(A_t^1 + A_t^2 - D_t^1 - D_t^2 > 0)dA_t^2 - I(q_t^2 > 0)dD_t^1 - I(q_t^2 > 0)dD_t^2, \quad (14)$$

т.е. очередь будет увеличивается на единицу, если в момент прихода новой заявки второго типа ($dA_t^2 = 1$) оператор занят, и уменьшаться на единицу, если в момент окончания обслуживания текущей заявки ($dD_t^1 = 1$ или $dD_t^2 = 1$) очередь заявок второго типа не пуста ($q_t^2 > 0$).

Логика построения уравнения (13) следующая. Во-первых, параметр t_t^{o2} принимает значение равное сумме текущего значения времени и параметра τ_2, если в момент прихода новой заявки второго типа ($dA_t^2 = 1$) оператор свободен, либо если в момент окончания обслуживания текущей заявки ($dD_t^1 = 1$ или $dD_t^2 = 1$) в очереди есть заявки второго типа ($q_t^2 > 0$). Во-вторых, обнуляется, если в момент окончания обслуживания заявки второго типа ($dD_t^2 = 1$) очередь заявок второго типа пуста ($q_t^2 = 0$).

Отметим, что уравнения построены так, что в любой момент времени $t \geq 0$ между параметрами t_t^{o1} и t_t^{o2} соблюдаются следующие соотношения. Во-первых, если $t_t^{o1} > 0$, то $t_t^{o2} = 0$. Во-вторых, если $t_t^{o2} > 0$, то $t_t^{o1} = 0$. Это связано с тем, что оператор не может одновременно обслуживать заявки двух типов.

Итерационные формулы

Выведем формулы, необходимые для имитационного моделирования СМО. На стохастическом базисе $B = (\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\geq 0}, P)$ из формул (1)-(14) можно получить следующие инфинитезимальные соотношения:

$$P\{A_{t+\Delta}^1 - A_t^1 = 1 | \mathcal{F}_t\} = \lambda_1 \cdot \Delta + o(\Delta), \quad (15)$$

$$P\{A_{t+\Delta}^2 - A_t^2 = 1 | \mathcal{F}_t\} = \lambda_2 \cdot \Delta + o(\Delta), \quad (16)$$

$$P\{D_{t+\Delta}^1 - D_t^1 = 1 | \mathcal{F}_t\} = \mu_t^1 \cdot \Delta + o(\Delta), \quad (17)$$

$$P\{D_{t+\Delta}^2 - D_t^2 = 1 | \mathcal{F}_t\} = \mu_t^2 \cdot \Delta + o(\Delta). \quad (18)$$

Основываясь на понятие геометрической вероятности, по формулам (15)-(18) проведем имитационное моделирование. А именно, введя дискретизацию (шаг по времени) Δ из условия $\lambda_1 \cdot \Delta \ll 1$, $\lambda_2 \cdot \Delta \ll 1$, $\mu_t^1 \cdot \Delta \ll 1$, $\mu_t^2 \cdot \Delta \ll 1$ получим следующие итерационные формулы (для вычисления значений процессов в момент времени $t + \Delta$ через значения процессов в момент t):

$$A_{t+\Delta}^1 = A_t^1 + \delta(\lambda_1), \quad (19)$$

$$A_{t+\Delta}^2 = A_t^2 + \delta(\lambda_2), \quad (20)$$

$$D_{t+\Delta}^1 = D_t^1 + \delta(\mu_t^1), \quad (21)$$

$$D_{t+\Delta}^2 = D_t^2 + \delta(\mu_t^2), \tag{22}$$

где $\delta(\gamma) = \begin{cases} 1, & \text{с вероятностью } \gamma \cdot \Delta, \\ 0, & \text{с вероятностью } 1 - \gamma \cdot \Delta. \end{cases}$

Оставшиеся параметры вычисляем следующим образом:

$$q_{t+\Delta}^1 = q_t^1 + I(A_t^1 + A_t^2 - D_t^1 - D_t^2 > 0)\Delta A_t^1 - I(q_t^1 > 0, q_t^2 = 0)\Delta D_t^1 - \\ -I(q_t^1 > 0, q_t^2 = 0)\Delta D_t^2, \tag{23}$$

$$t_{t+\Delta}^{o1} = t_t^{o1} + (t + \tau_1) \cdot I(A_t^1 + A_t^2 - D_t^1 - D_t^2 = 0)\Delta A_t^1 + \\ +(t + \tau_1 - t_t^{o1}) \cdot I(q_t^1 > 0, q_t^2 = 0)\Delta D_t^1 + (t + \tau_1) \cdot I(q_t^1 > 0, q_t^2 = 0)\Delta D_t^2 - \\ -t_t^{o1} I(q_t^1 = 0)\Delta D_t^1 - t_t^{o1} - t_t^{o1} I(q_t^2 > 0)\Delta D_t^1, \tag{24}$$

$$q_{t+\Delta}^2 = q_t^2 + I(A_t^1 + A_t^2 - D_t^1 - D_t^2 > 0)\Delta A_t^2 - I(q_t^2 > 0)\Delta D_t^1 - I(q_t^2 > 0)\Delta D_t^2, \tag{25}$$

$$t_{t+\Delta}^{o2} = t_t^{o2} + (t + \tau_2) \cdot I(A_t^1 + A_t^2 - D_t^1 - D_t^2 > 0)\Delta A_t^2 + (t + \tau_2) \cdot I(q_t^2 > 0)\Delta D_t^1 + \\ +(t + \tau_2) \cdot I(q_t^2 > 0)\Delta D_t^2 - t_t^{o2} I(, q_t^2 = 0)\Delta D_t^2. \tag{26}$$

Здесь $\Delta A_t^1 = A_{t+\Delta}^1 - A_t^1,\ \Delta A_t^2 = A_{t+\Delta}^2 - A_t^2,\ \Delta D_t^1 = D_{t+\Delta}^1 - D_t^1, \Delta D_t^2 = D_{t+\Delta}^2 - D_t^2$.

Результаты компьютерного моделирования

Практическая реализация СМО осуществлена с помощью языка программирования высокого уровня C# в среде разработки Visual Studio 2022. На рисунке 2 представлен результат моделирования систем при параметрах $\tau_1 = 2,\ \tau_2 = 1,\ \lambda_1 = 2,\ \lambda_2 = 1$ и времени моделирования $T = 10$.

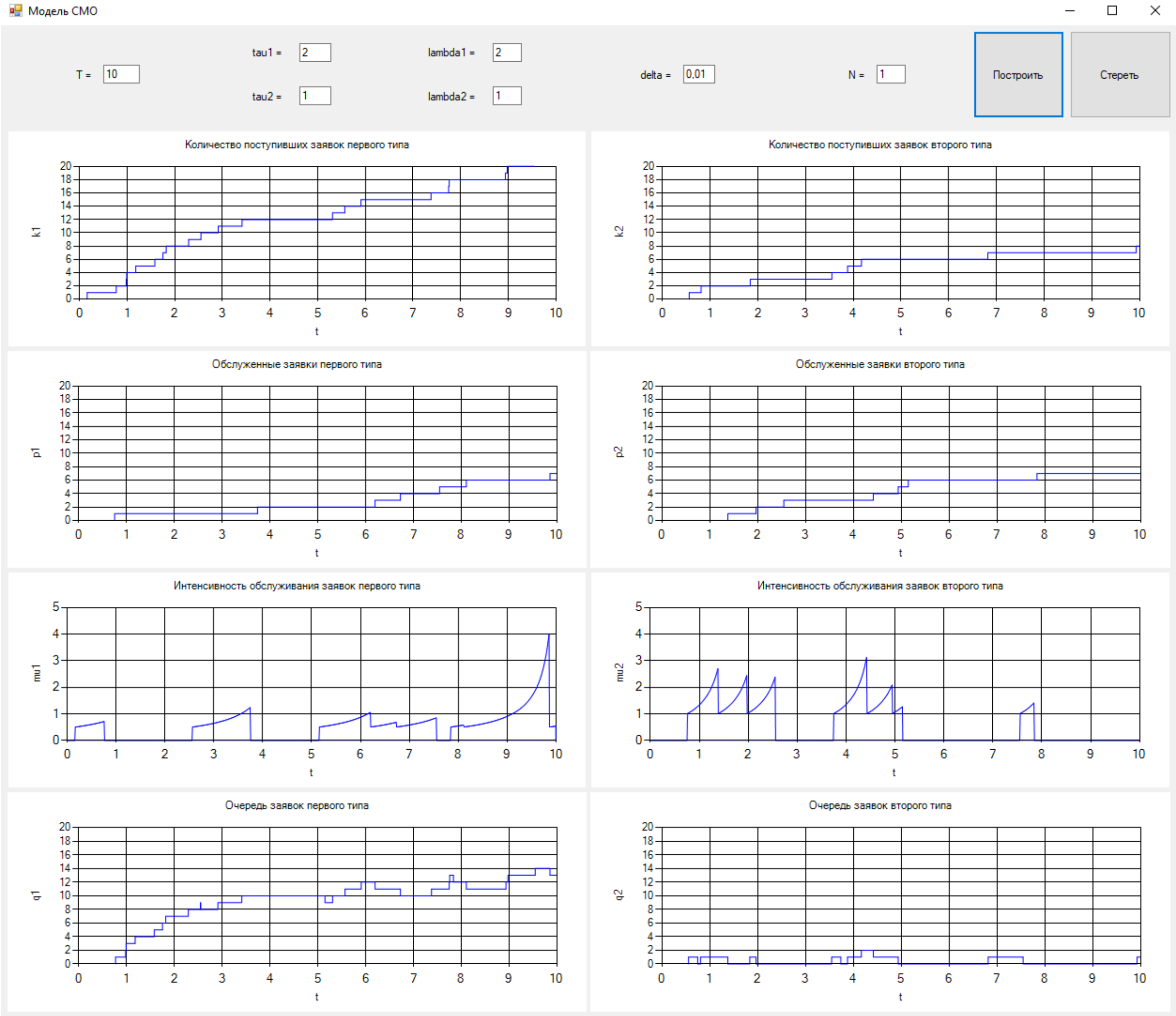


Рисунок 2 - Модель СМО

Результаты моделирования показывают, что с данными входными параметрами система перегружена, растет очередь заявок первого типа. Однако, важные заявки обслуживаются своевременно, что и являлось основной задачей модели.

Заключение

В результате выполнения данной работы была построена математическая модель системы массового обслуживания «точно-в-срок» с относительным приоритетом в семимартингальных терминах. Показан переход от математической модели к итерационным формулам, по которым было проведено имитационное моделирование.

Список литературы

1. Butov A.A., Kovalenko A.A. Stochastic models of simple controlled systems just-in-time // Вестник Самарского государственного технического университета. Серия: Физикоматематические науки. 2018, т. 22, №. 3, с. 518-531.
2. Бутов А.А. Оценивание параметров распределенных продуктивных систем, работающих по принципу «точно в срок» // Автомат. и телемех. 2020, № 3, с.14–27.

3. Бородин А.Н. Случайные процессы: Учебник. Спб.: Изд-во «Лань», 2013.
4. Бутов, А.А. Теория случайных процессов и её дополнительные главы: учеб. пособие. Ч. 1. Введение в стохастическое исчисление. Ульяновск : УлГУ, 2016
5. Бутов, А.А. Теория случайных процессов и её дополнительные главы: учеб. пособие. Ч. 2. Случайное блуждание, винеровский процесс, стохастический интеграл, диффузионные процессы. Ульяновск : УлГУ, 2021

References

1. Butov A.A., Kovalenko A.A. Stochastic models of simple controlled systems just-in-time // Bulletin of the Samara State Technical University. Series: Physical and Mathematical Sciences. 2018, vol. 22, No. 3, pp. 518-531.
2. Butov A.A. Estimation of parameters of distributed productive systems operating on the principle of "just in time" // Automaton. and telemech. 2020, No. 3, pp.14-27.
3. Borodin A.N. Random processes: Textbook. St. Petersburg: Publishing house "Lan", 2013.
4. Butov, A.A. Theory of Random Processes and Its Additional Chapters. allowance. Part 1. Introduction to Stochastic Calculus. Ulyanovsk : Ulyanovsk State University, 2016
5. Butov, A.A. Theory of Random Processes and Its Additional Chapters. allowance. Part 2. Random walk, Wiener process, stochastic integral, diffusion processes. Ulyanovsk : Ulyanovsk State University, 2021