Об одной обобщённой системе Эрланга—Севастьянова и её скорости сходимости

Г. А. ЗВЕРКИНА
Российский университет транспорта (МИИТ),
Институт проблем управления им. В. А. Трапезникова РАН
e-mail: zverkina@gmail.com

УДК 519.21

Ключевые слова: распределение состояния системы массового обслуживания, скорость сходимости, регенерирующие процессы, метод склеивания.

Аннотация

Получена оценка сверху для скорости сходимости распределения состояния бесконечнолинейной системы массового обслуживания в случае, когда интенсивность входящего потока и интенсивность обслуживания зависят от состояния системы.

Abstract

G. A. Zverkina, On some extended Erlang–Sevastyanov queueing system and its convergence rate, Fundamentalnaya i prikladnaya matematika, vol. 22 (2018), no. 3, pp. 57—82.

An upper bound for the convergence rate of the distribution of a queuing system state with infinitely many servers is obtained for the case where the intensities of the incoming and service flows depend on the state of the system.

1. Введение

1.1. Мотивация

Рассматривается бесконечнолинейная система массового обслуживания (СМО), в которую в моменты времени $t_1, t_2, \ldots, t_n$ поступают заявки, которые сразу же начинают обслуживаться (полагаем $t_0 \stackrel{\text{def}}{=} 0$). Время обслуживания i-й заявки — случайная величина ξ_i с функцией распределения $G_i(s)$, а промежутки между поступлениями требований $\eta_i \stackrel{\text{def}}{=} t_i - t_{i-1}$ — случайная величина с функцией распределения $F_i(s)$. Для простоты предположим, что распределения случайных величин ξ_i и η_i абсолютно непрерывны; обозначим $f_i(s) \stackrel{\text{def}}{=} F_i'(s)$, $g_i(s) \stackrel{\text{def}}{=} G_i'(s)$.

Мы будем описывать распределения $F_i(s)$ и $G_i(s)$ в терминах *интенсивностей*.

Фундаментальная и прикладная математика, 2018, том 22, № 3, с. 57—82.
© 2018 *Национальный Открытый Университет «ИНТУИТ»*

Определение 1. *Интенсивностью* в момент времени $s > 0$ окончания временно́го периода $\mathscr{T}$, начавшегося в момент времени $t = 0$ и не завершившегося до момента времени $s > 0$, называется величина

$$\phi(s) \stackrel{\text{def}}{=} \lim_{\Delta \downarrow 0} \frac{\mathsf{P}\{\mathscr{T} \in (s, s+\Delta] \mid \mathscr{T} > s\}}{\Delta}.$$

Если $\mathscr{T}$ имеет абсолютно непрерывную функцию распределения $\Phi(x)$, то

$$\phi(s) = \lim_{\Delta \downarrow 0} \frac{\Phi(s+\Delta) - \Phi(s)}{\Delta\bigl(1 - \Phi(s)\bigr)} = \frac{\Phi'(s+0)}{1 - \Phi(s)}.$$

Замечание 1. Из определения 1 следует, что

$$\mathsf{P}\{\mathscr{T} \in (s, s+\Delta] \mid \mathscr{T} > s\} = \phi(s)\Delta + o(\Delta).$$

Замечание 2. Функция распределения и её плотность легко восстанавливаются по интенсивности:

$$\Phi(s) = 1 - \exp\biggl(-\int_0^s \phi(u)\,\mathrm{d}u\biggr); \tag{1}$$

$$\Phi'(s) = \phi(s)\exp\biggl(-\int_0^s \phi(u)\,\mathrm{d}u\biggr). \tag{2}$$

Очевидно, постоянная интенсивность соответствует экспоненциальному распределению. Интенсивность можно также рассматривать в случае, когда функция распределения $\Phi(s)$ имеет разрывы. Но в этом случае формулы (1) и (2) неприменимы.

Итак, мы описываем поведение бесконечнолинейной СМО на языке интенсивностей.

1. Интенсивность поступления очередной заявки в момент времени t есть $\lambda(t)$, т. е.

$$\mathsf{P}\left\{\begin{matrix}\text{Очередная заявка поступит} \\ \text{в интервале } (t, t+\Delta]\end{matrix}\right\} = \lambda(t)\Delta + o(\Delta).$$

2. Интенсивность обслуживания i-й по порядку поступления заявки из имеющихся n заявок в системе в момент времени t равна $h_i(t)$, т. е.

$$\mathsf{P}\left\{\begin{matrix}i\text{-е требование,} \\ \text{обслуживающееся} \\ \text{в момент времени } t, \\ \text{закончит обслуживание} \\ \text{на промежутке } (t, t+\Delta]\end{matrix}\right\} = \mathsf{P}\{\xi_i \in (s, s+\Delta] \mid \xi_i > s\} = h_i(t)\Delta + o(\Delta).$$

Здесь и везде далее под i-м требованием, находящимся в СМО, мы понимаем i-е по порядку поступления требование из находящихся в СМО.

В случае когда $\lambda(t) \equiv \lambda$ и $h_i(t) \equiv \mu$, мы имеем дело с СМО $M|M|\infty$, впервые изученной в [10]; было показано, что при условии $\mu < \lambda$ распределение количества заявок в системе n_t сходится к стационарному распределению:

$$\lim_{t\to\infty} \mathsf{P}\{n_t = k\} = \mathscr{P}_k = \mathscr{P}_0 \frac{\lambda^k}{\mu^k},$$

где

$$\mathscr{P}_0 = 1 - \frac{\lambda}{\mu}.$$

Позднее этот результат был обобщён на случай, когда интенсивность входящего потока зависит только от числа требований в системе, т. е. $\lambda(t) \equiv \lambda_k$ при $n_t = k$ (см., например, [15, гл. 4, § 4, 5]):

$$\lim_{t\to\infty} \mathsf{P}\{n_t = k\} = \mathscr{P}_k = \mathscr{P}_0 \frac{\prod\limits_{i=0}^{k-1} \lambda_i}{\mu^k}, \quad \text{где} \quad \mathscr{P}_0 = \left(\sum_{k=0}^{\infty} \frac{\prod\limits_{i=0}^{k-1} \lambda_i}{\mu^k} \right)^{-1}. \tag{3}$$

Более того, если интенсивность обслуживания всех находящихся в СМО заявок зависит только от количества заявок, т. е. $h_i(t) \equiv \mu_k > 0$ при $n_t = k$, формула 3 применима, но с заменой μ^k на $\prod\limits_{i=1}^{k} \mu_i$. Очевидно, если $\lambda_k = 0$ при некотором k, то (3) применима для системы с конечным числом серверов без очереди.

В описанном случае, когда

$$\begin{gathered} \lambda(t) \not\equiv \lambda, \quad \mu(t) \not\equiv \mu, \\ \lambda(t) = \lambda_{n_t}, \quad h_i(t) = \mu_{n_t}, \end{gathered} \tag{4}$$

где n_t — количество требований в системе в момент времени t, времена обслуживания и входящий поток зависимы между собой.

В [11] исследовалась СМО с пуассоновским входящим потоком и произвольным абсолютно непрерывным распределением времени обслуживания (СМО $M|G|\infty$), одинаковым для всех требований; предполагалось, что времена обслуживания и входящий поток независимы. В терминах интенсивностей это означает, что если в момент времени t в системе находится n_t требований и i-е требование находилось на обслуживании в течение времени $x_t^{(i)}$ $(i = 1, 2, \ldots, n)$, то интенсивность обслуживания i-го требования в момент времени t равна $h_i(t) = h\big(x_t^{(i)}\big)$.

Обозначим

$$\mu \stackrel{\text{def}}{=} \left(\int\limits_0^{\infty} s \, \mathrm{d}G(s) \right)^{-1},$$

где

$$G(s) = 1 - \exp\left(- \int\limits_0^{s} h(u) \, \mathrm{d}u \right) -$$

функция распределения времени обслуживания. В [11] было сформулировано утверждение, что для $k > 0$

$$\lim_{t\to\infty} \mathsf{P}\{n_t = k;\ x_t^{(1)} < \alpha_1,\ x_t^{(2)} < \alpha_2, \dots,\ x_t^{(k)} < \alpha_k\} = \\ = \mathscr{P}(k; \alpha_1, \alpha_2, \dots, \alpha_k) = \mathscr{P}_0 \prod_{i=1}^{k} \lambda_i G(1-\alpha_i),$$

где нормирующий множитель $\mathscr{P}_0$ вычисляется так же, как в (3).

Этот факт для СМО $M|G|m|0$ в более общем виде, без предположения об абсолютной непрерывности функции распределения времени обслуживания $G(s)$, был доказан в [4] (см. также [5]).

Таким образом, в [11] было предложено рассматривать СМО, в которой

$$\lambda(t) \equiv \lambda, \quad h_i(t) = h_i\big(x_t^{(i)}\big), \tag{5}$$

где $x_t^{(i)}$ — прошедшее к моменту времени t время обслуживания i-го по порядку поступления требования из n_t требований, находящихся в системе.

В дальнейшем результаты, основанные на [10], были получены многими исследователями (см., например, [1, 8, 14, 18]).

Естественным обобщением ситуаций (4) и (5) является СМО, в которой интенсивности входящего потока и обслуживания зависят не только от n_t и прошедших времён обслуживания находящихся на обслуживании требований, но и от того, когда произошло последнее изменение n_t — количества требований в системе:

$$\lambda(t) = \lambda\big(x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(n_t)}\big), \quad h_i(t) = h_i\big(x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(n_t)}\big),$$

где $x_t^{(0)}$ — время, прошедшее с последнего изменения количества требований в системе n_t, а $x_t^{(i)}$ — прошедшее к моменту времени t время обслуживания i-го по порядку поступления требования из n_t требований, находящихся в системе. Знание стационарного распределения состояния СМО чрезвычайно важно для приложений теории массового обслуживания, на основании этих распределений рассчитываются технические характеристики сложных систем. Такого рода СМО были рассмотрены, например, в [21, 22, 24].

Например, в [24] был доказан следующий факт: для СМО с одним обслуживающим прибором и бесконечной очередью в случае, когда интенсивность обслуживания h и интенсивность входящего потока λ зависят от количества заявок в системе n_t и прошедшего времени обслуживания x_t, верна следующая теорема.

Теорема 1. *Обозначим $X_t \stackrel{\text{def}}{=} (n_t, x_t)$, где n_t — число находящихся в СМО требований в момент времени t, x_t — время, в течение которого находилось на обслуживании требование, которое в момент времени t находится на обслуживании. В случае $n_t = 0$ интенсивность входящего потока постоянна и равна $\lambda_0 > 0$ и $X_t \stackrel{\text{def}}{=} (0, 0)$.*

Если $\lambda(n_t, x_t)$ и $h(n_t, x_t)$ измеримы по Борелю и ограниченны, выполнены условия

$$\sup_{n,x\colon n>0} \lambda(n,x) = \Lambda < \infty, \qquad \inf_{n>0,\ x\geqslant 0} h(n,x) \geqslant \frac{C_0}{1+x},$$

$$C_0 > 4(1+2\Lambda) \text{ и для некоторого } k>1 \quad C_0 > 2^{k+1}(1+\Lambda 2^k),$$

то распределение процесса $X_t = (n_t, x_t)$ сходится к стационарному распределению $\mathscr{P}$ на пространстве состояний

$$\mathscr{X} = \{(0,0)\} \cup \{(n,x)\colon n \in \mathbb{N},\ x \geqslant 0\}$$

и найдутся такие $m > k$ и $C > 0$, что при всех $t \geqslant 0$

$$\sup_{S\in\mathscr{B}(\mathscr{X})} |\mathsf{P}\{X_t \in S \mid X_0 = (n,x)\} - \mathscr{P}(S)| \leqslant C\,\frac{(1+n+x)^m}{(1+x)^k},$$

где $\mathscr{P}$ — стационарное распределение процесса X_t.

В доказательстве теоремы 1 приводится (неполный) алгоритм вычисления константы C.

Однако существует достаточно большое количество результатов о *порядке* скорости сходимости распределения состояния СМО к стационарному распределению (см., например, [1, 3, 7, 19] и др.). Доказано, что некоторые характеристики СМО χ_t сходятся к стационарному значению $\tilde{\chi}$ со скоростью $\phi(t)$, т. е. существует такая константа K, что при всех t выполняется $|\chi_t - \tilde{\chi}| < K\phi(t)$, где $\phi(t) \searrow 0$.

Так, например, в СМО, исследованной в [24] (теорема 1), для сходимости распределения процесса X_t к стационарному распределению со скоростью, не превосходящей $K/(1+x)^k$, $k \in [0, C_0 - 1)$, достаточно $\Lambda > C_0$, но без дополнительных условий теоремы 1 оценить константу K ранее не удавалось.

Отметим, что для СМО $MI|GI|\infty$ хорошо известны (см., например, [18]) такие характеристики, как $\mathscr{P}_k(t)$ — вероятность того, что в системе занято k приборов в момент времени t, известны средняя длина периода занятости и его дисперсия, а также преобразование Лапласа функции распределения периода занятости (если в начальный момент времени система свободна). Эти результаты позволяют легко оценить скорость сходимости величин $\mathscr{P}_k(t)$ к их стационарным значениям $\mathscr{P}_k$ (см., например, [12]).

1.2. Известные факты и цель работы

Напомним, что СМО $MI|GI|\infty$ состоит из бесконечного числа приборов, работающих независимо, и случайное время обслуживания попавшего на прибор требования (заявки) одинаково распределено для всех приборов и имеет функцию распределения $G(s)$; входящий поток пуассоновский с постоянным параметром Λ. Предполагается, что в начальный момент времени система свободна (нет требований на обслуживании).

Цель настоящей работы — исследование поведения бесконечнолинейной СМО, в которой входящий поток и времена обслуживания задаются интенсивностями, зависящими от *полного состояния системы*; при этом все времена обслуживания находящихся в системе требований *некоторым образом* зависят друг от друга, а также они зависят и от входящего потока требований, который может менять свою интенсивность в зависимости от *полного состояния системы*. При этом естественно предполагать, что, несмотря на такую зависимость, некоторые характеристики времени обслуживания и входящего потока находятся в определённых границах (например, естественно требовать, чтобы интенсивность входящего потока была ограничена, а интенсивность обслуживания не обращалась в нуль).

В работе исследуется ситуация, когда интенсивность входящего потока, а также интенсивность обслуживания зависят от *полного состояния системы*, а именно от того времени, которое прошло с момента поступления последнего пришедшего требования, и от того времени, которое имеющиеся в системе требования находились на обслуживании.

Как уже говорилось, несмотря на наличие указанной зависимости случайных величин (времена между поступлениями требований и времена обслуживания требований), мы предполагаем некоторые ограничения для интенсивности поступления требований в систему и интенсивностей их обслуживания.

Предложенный выше вариант зависимости времён обслуживания и входящего потока позволяют утверждать, что процесс обслуживания такой системы марковский и *регенерирующий*. Точками регенерации такого процесса являются моменты времени, когда все входящие в *полное состояние системы* величины обнуляются. Это может произойти тогда, когда в свободную систему приходит требование, т. е. полное состояние системы содержит нулевое время обслуживания единственного требования и нулевое время ожидания поступления следующего требования.

2. Описание изучаемой системы массового обслуживания

Итак, мы рассматриваем СМО, в которой интенсивность входящего потока и интенсивность обслуживания зависят от полного состояния системы, описание которого будет дано ниже.

Полное состояние системы в момент времени t включает время $x_t^{(0)}$, прошедшее с момента прихода последней заявки, и времена $x_t^{(i)}$, прошедшие с момента начала обслуживания i-го требования из общего количества находящихся (требования нумеруются в соответствии с порядком их поступления), также для удобства мы будем использовать переменную n_t — число заявок в системе в момент времени t. Таким образом, полное состояние системы описывается

вектором $X_t = \big(n_t, x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\big)$; поскольку i-я заявка поступила ранее $(i+1)$-й, ясно, что $x_t^{(i)} \geqslant x_t^{(i+1)}$ для $i = 1, \ldots, n_t$. Легко заметить, что $x_t^{(0)} \leqslant x_t^{(n_t)}$, т. е. заявка с номером n_t появилась в системе не позднее времени $t - x_t^{(0)}$.

Мы предполагаем, что интенсивность входящего потока есть функция от X_t: $\lambda = \lambda(X_t)$, как интенсивность обслуживания i-й заявки (которая уже обслуживалась в течение времени $x_t^{(i)}$) — это функция $h_i = h_i(X_t)$. Это значит, что для малого промежутка времени $\Delta > 0$ при заданном $X_t = \big(n_t, x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\big)$ — напомним, что $x_t^{(i)} \geqslant x_t^{(i+1)}$ — верно следующее:

$$\left.\begin{aligned}
&\mathsf{P}\left\{\begin{array}{l} n_{t+\Delta} = n_t + 1,\ x_{t+\Delta}^{(0)} = x_{t+\Delta}^{(n_{t+\Delta})} \in (0;\Delta) \\ \text{и } x_{t+\Delta}^{(i)} = x_t^{(i)} + \Delta \text{ для всех } i = 1, \ldots, n_t, \\ x_{t+\Delta}^{(n_t+1)} \in (0;\Delta) \end{array}\right\} = \lambda(X_t)\Delta + o(\Delta); \\
&\mathsf{P}\{n_{t+\Delta} = n_t - 1\} = \sum_{i=1}^{n_t} h_i(X_t)\Delta + o(\Delta); \\
&\mathsf{P}\left\{\begin{array}{l} n_{t+\Delta} = n_t - 1 \text{ и} \\ 1) \text{ для всех } j < i\ x_{t+\Delta}^{(j)} = x_t^{(j)} + \Delta, \\ 2) \text{ для всех } j > i\ x_{t+\Delta}^{(j)} = x_t^{(j+1)} + \Delta \end{array}\right\} = h_i(X_t)\Delta + o(\Delta) \\
&\quad \text{для всех } i = 1, 2, \ldots, n_t; \\
&\mathsf{P}\big\{n_{t+\Delta} = n_t;\ x_{t+\Delta}^{(i)} = x_t^{(i)} + \Delta \text{ для всех } i = 0, \ldots, n_t\big\} = \\
&\quad = 1 - \left(\lambda(X_t) + \sum_{i=1}^{n_t} h_i(X_t)\right)\Delta + o(\Delta).
\end{aligned}\right\} \qquad (6)$$

Из (6) следует, что распределение $X_{t+\Delta}$ зависит только от распределения X_t, т. е. X_t — марковский процесс на пространстве состояний

$$\mathscr{X} \stackrel{\text{def}}{=} \bigcup_{i=0}^{\infty} \mathscr{S}_i,$$

где

$$\mathscr{S}_i \stackrel{\text{def}}{=} \{i\} \times \prod_{j=0}^{i} \mathbb{R}_+, \quad i \in \mathbb{Z}_+,$$

множество $\mathscr{S}_0$ — это множество свободных состояний СМО (когда заявки в системе отсутствуют).

Обозначаем через $\mathscr{P}_t^{X_0}$ распределение процесса X_t с начальным состоянием X_0:

$$\mathscr{P}_t^x(S) = \mathsf{P}\{X_t \in S \mid X_0 = x\}, \quad x \in \mathscr{X}, \quad S \in \mathscr{B}(\mathscr{X}).$$

Как уже говорилось, моменты попадания процесса X_t в состояние $(1,0;0)$, т. е. выхода процесса X_t из множества $\mathscr{S}_0$ (в свободную систему поступило требование), являются моментами регенерации процесса X_t. Поэтому если время между моментами регенерации имеет конечное математическое ожидание, то процесс X_t эргодический и его распределение слабо сходится к некоторому стационарному (инвариантному) распределению.

Замечание 3. Предложенным способом может быть описан широкий класс СМО, например рассмотренные выше СМО $M|M|\infty$, $M|G|\infty$, $M|G|m|0$ и другие.

Для того чтобы среднее периода регенерации хорошо изученной СМО $M|G|\infty$ было конечным, достаточно, чтобы среднее времени обслуживания было конечным. При этом скорость сходимости величин $\mathscr{P}_k(t)$ к их стационарным значениям $\mathscr{P}_k$ тем выше, чем больше конечных моментов у распределения времени обслуживания (см., например, [12]).

Естественно поэтому рассматривать ситуацию, когда (зависимые) времена обслуживания имеют конечные моменты, а характеристики входящего потока сравнимы с характеристиками пуассоновского потока.

Наложим на интенсивности следующие условия:

$$h_i(X_t) \geqslant \frac{K}{1+x_t^{(i)}}, \quad K>2; \quad 0<\lambda_0 \leqslant \lambda(X_t) \leqslant \Lambda < \infty. \tag{7}$$

Из первого условия следует, что функция распределения времени обслуживания $G_i(s)$ удовлетворяет неравенству

$$G_i(s) \geqslant 1 - \frac{1}{(1+s)^K},$$

т. е. время обслуживания имеет $k \in [0;K)$ конечных моментов (см. замечание 2). Также плотность функции распределения интервалов времени поступлениями заявок $F(s)$ удовлетворяет неравенству

$$f_i(s) \leqslant \Lambda e^{-\lambda_0 s},$$

и распределение времён η_i между поступлениями заявок имеет любые степенны́е моменты:

$$\mathbb{E}\,(\eta_i)^k \leqslant \frac{k!\,\Lambda}{\lambda_0^{k+1}}.$$

В частности, эта оценка верна для *свободных периодов* X_t, т. е. промежутков времени, где $n_t \equiv 0$.

Наша цель — найти вычисляемые оценки скорости сходимости распределения процесса X_t к стационарному распределению в условиях (7). Для этого мы воспользуемся *методом склеивания* (см. раздел 5). Однако сначала мы напомним некоторые сведения из теории массового обслуживания и получим необходимые в дальнейшем оценки.

3. Некоторые сведения о системе $MI|GI|\infty$

Здесь мы приведём известные результаты о системе $MI|GI|\infty$ (см. [17, 18] и др.).

В систему $MI|GI|\infty$ поступает пуассоновский поток заявок с постоянной интенсивностью Λ. Пришедшая заявка сразу начинает обслуживание. Её время обслуживания является случайной величиной (обозначаемой ξ_i) с функцией распределения $G(s)$. Все случайные величины (времена обслуживания и времена между поступлениями заявок) независимы в совокупности.

В начальный момент времени СМО свободна.

Предполагается, что существует некоторое количество конечных моментов случайной величины X_i:

$$\mathbb{E}\,(\xi_i)^k = k\int\limits_0^\infty s^{k-1}\big(1-G(s)\big)\,\mathrm{d}s = m_k < \infty$$

В дальнейшем положим

$$G(s) = 1 - \frac{1}{(1+s)^K}, \quad K > 2. \tag{8}$$

Условие (8) гарантирует наличие $K-\varepsilon$ ($\varepsilon > 0$) конечных моментов; очевидно, что

$$\mathbb{E}\,\xi_i = \frac{1}{K-1}, \quad \mathbb{E}\,(\xi_i)^2 = \frac{2}{K^2-3K+2} \quad \text{и т. д.}$$

Нагрузкой СМО назовём величину $\rho \stackrel{\text{def}}{=} \Lambda m_1$.

Опять описываем поведение процесса обслуживания марковским процессом

$$\overset{\circ}{X}_t = \Big(\overset{\circ}{n}_t, \overset{\circ}{x}{}_t^{(0)}; \overset{\circ}{x}{}_t^{(1)}, \overset{\circ}{x}{}_t^{(2)}, \ldots, \overset{\circ}{x}{}_t^{(\overset{\circ}{n}_t)}\Big),$$

где $\overset{\circ}{n}_t$ — число требований в СМО в момент времени t, а $\overset{\circ}{x}{}_t^{(0)}$ — время, прошедшее с момента поступления последнего требования; $\overset{\circ}{x}{}_t^{(i)}$ — время, в течение которого i-я по порядку поступления заявка из находящихся в системе ($i = 1, 2, \ldots, \overset{\circ}{n}_t$) находилась на обслуживании. (Здесь и везде в дальнейшем знак $\circ$ над случайной величиной, процессом или их числовыми характеристиками обозначает, что речь идёт о «классической» СМО $MI|GI|\infty$.)

Распределение процесса $\overset{\circ}{X}_t$ обозначаем $\overset{\circ}{\mathscr{P}}_t$: $\overset{\circ}{\mathscr{P}}_t(S) = \mathsf{P}\{\overset{\circ}{X}_t \in S\}$ для всех $S \in \mathscr{B}(\mathscr{X})$, и снова

$$\mathscr{X} \stackrel{\text{def}}{=} \bigcup_{i=0}^{\infty} \mathscr{S}_i$$

и

$$\mathscr{S}_i \stackrel{\text{def}}{=} \{i\} \times \prod_{j=0}^{i} \mathbb{R}_+, \quad i \in \mathbb{Z}_+.$$

В этих предположениях известно (см., например, [18]), что

$$\mathcal{P}_k(t) \stackrel{\text{def}}{=} \mathsf{P}\{\overset{\circ}{n}_t = k\} = \overset{\circ}{\mathscr{P}}_t\,(\mathscr{S}_0) = \frac{e^{-\mathfrak{G}(t)}\big(\mathfrak{G}(t)\big)^k}{k!},$$

где

$$\mathfrak{G}(t) \stackrel{\text{def}}{=} \Lambda \int\limits_0^t \big(1 - G(s)\big)\,\mathrm{d}s.$$

Замечание 4. $\mathfrak{G}(+\infty) = \Lambda m_1$, и $\mathcal{P}_0(t) \geqslant e^{-\Lambda m_1} = e^{-\rho}$ при $t \geqslant 0$.

Кроме этого, известно распределение периода занятости СМО $M|G|\infty$ (см., например, [17]). Обозначим через $\overset{\circ}{\zeta}_i$ i-й период занятости СМО $M|G|\infty$. Эта случайная величина имеет функцию распределения

$$B(x) \stackrel{\text{def}}{=} \mathsf{P}\{\overset{\circ}{\zeta}_i \leqslant x\} = 1 - \frac{1}{\Lambda}\sum_{k=1}^{\infty} c^{n*}(x),$$

где

$$c(x) \stackrel{\text{def}}{=} \Lambda\big(1 - G(x)\big)e^{-\mathfrak{G}(x)},$$

а c^{n*} — это n-я свёртка функций $c(x)$.

Также можно выписать преобразование Лапласа для $B(x)$ (см., например, [17]):

$$\mathcal{L}[B](s) = 1 + \frac{s}{\Lambda} - \frac{1}{\Lambda \int\limits_0^\infty \exp\Big(-st - \Lambda \int\limits_0^t [1 - G(v)]\,\mathrm{d}v\Big)\,\mathrm{d}t}. \tag{9}$$

Впрочем, моменты периода занятости можно вычислять и по формулам

$$\mathbb{E}\,(\overset{\circ}{\zeta}_i)^n = (-1)^{n+1}\left\{\frac{e^\rho}{\Lambda} n C^{(n-1)} - e^\rho \sum_{k=1}^{n-1} C_n^k \mathbb{E}\,(\overset{\circ}{\zeta}_i)^{n-k} C^{(k)}\right\}, \quad n \in \mathbb{N}, \tag{10}$$

где

$$C^{(k)} = \int\limits_0^\infty (-t)^k \left(\exp\left(-\Lambda \int\limits_0^\infty [1 - G_0(v)]\,\mathrm{d}v\right)\right)\Lambda[1 - G_0(t)]\,\mathrm{d}t.$$

Используя (9) и (10), можно получить значения

$$\begin{aligned} &\mathbb{E}\,\overset{\circ}{\zeta}_i = \frac{e^\rho - 1}{\Lambda}, \\ &\mathbb{E}\,(\overset{\circ}{\zeta}_i)^2 = \frac{2e^{2\rho}}{\Lambda} \int\limits_0^\infty \left(\exp\left(-\Lambda \int\limits_0^\infty [1 - G(v)]\,\mathrm{d}v\right)\right)\mathrm{d}t + \frac{2e^\rho}{\Lambda^2} - \frac{2}{\Lambda^2} - 2\frac{e^\rho - 1}{\Lambda^2}. \end{aligned} \tag{11}$$

Но для более высоких моментов ζ_i вычисления по формулам (9) и (10) очень сложны и трудно применимы на практике.

Однако нам в дальнейшем потребуются оценки моментов ζ_i. Для этого используем следующее утверждение.

Утверждение 1. *Если* $\mathbb{E}\,\xi_i^k < \infty$ *при некотором* $k \in \mathbb{N}$, *то верна оценка*

$$\mathbb{E}\,(\overset{\circ}{\zeta}_i)^k \leqslant \frac{\mathbb{E}\,\xi_i^k}{1-e^{-\rho}} \times \varphi\big((1-e^{-\rho}), k-1\big), \tag{12}$$

где

$$\varphi(x,k) \overset{\text{def}}{=} \sum_{n=1}^{\infty} n^{k-1}x^n = \left(x\,\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \frac{1}{1-x}.$$

Доказательство. Функция $c(x)$, используемая в (9), неотрицательна, и

$$\int\limits_0^\infty c(x)\,\mathrm{d}x = \int\limits_0^\infty \Lambda\big(1-G(x)\big)\exp\left(-\Lambda\int\limits_0^x \big(1-G(s)\big)\,\mathrm{d}s\right)\mathrm{d}x =$$

$$= \left(1-\exp\left(-\Lambda\int\limits_0^\infty \big(1-G(s)\big)\,\mathrm{d}s\right)\right) = \big(1-\exp(-\Lambda m_1)\big) = \big(1-\exp(-\rho)\big) \overset{\text{def}}{=} \varrho,$$

где опять $\rho = \Lambda m_1$. Таким образом, функция $\varsigma(x) \overset{\text{def}}{=} \varrho^{-1}\,c(x)$ является плотностью распределения некой случайной величины ϑ, а $\varsigma^{n*}(x)$ — это плотность распределения $\sum\limits_{j=1}^{n} \vartheta_j$ — суммы n одинаково распределённых и независимых случайных величин с тем же распределением, что и ϑ. Поэтому

$$\int\limits_0^\infty \varsigma^{n*}(x)\,\mathrm{d}\,x = 1.$$

Следовательно,

$$\mathbb{E}\;\overset{\circ}{\zeta}_i = \int\limits_0^\infty \big(1-B(x)\big)\,\mathrm{d}x = \sum_{n=1}^{\infty}\frac{1}{\Lambda}\int\limits_0^\infty c^{n*}(x)\,\mathrm{d}x =$$

$$= \frac{1}{\Lambda}\sum_{n=1}^{\infty}\varrho^n\int\limits_0^\infty \varsigma^{n*}(x)\,\mathrm{d}x = \frac{1}{\Lambda}\sum_{n=1}^{\infty}\varrho^n = \frac{1}{\Lambda}\times\frac{\varrho}{1-\varrho} = \frac{\exp(\rho)-1}{\Lambda}.$$

Далее, если $m_{k+1} = \mathbb{E}\,\xi_i^{k+1} < \infty$ при некотором $k \in \mathbb{N}$, то

$$\mathbb{E}\,\vartheta^k = \int\limits_0^\infty x^k\varsigma(x)\,\mathrm{d}\,x = \int\limits_0^\infty x^k\varrho^{-1}\Lambda\big(1-G(x)\big)e^{-\mathfrak{G}(x)}\,\mathrm{d}\,x \leqslant$$

$$\leqslant \frac{\Lambda}{\varrho}\int\limits_0^\infty x^k\big(1-G(x)\big)\,\mathrm{d}\,x = \frac{\Lambda\mathbb{E}\,\xi_i^{k+1}}{\varrho(k+1)}$$

и

$$\int\limits_0^\infty x^k\varsigma^{n*}(x)\,\mathrm{d}\,x = \mathbb{E}\left(\sum_{j=1}^{n}\vartheta_j\right)^k \leqslant n^k\mathbb{E}\,\vartheta^k \leqslant \frac{n^k\Lambda m_{k+1}}{\varrho(k+1)}.$$

В последней оценке мы использовали неравенство Йенсена (для $k \geqslant 1$) в виде

$$(a_1 + \ldots + a_n)^k \leqslant n^{k-1}(a_1^k + \ldots + a_n^k). \tag{13}$$

Теперь оценим $\mathbb{E}\,(\overset{\circ}{\zeta}_i)^k$:

$$\mathbb{E}\,(\overset{\circ}{\zeta}_i)^k = \int\limits_0^\infty kx^{k-1}\big(1 - B(x)\big)\,\mathrm{d}x = \frac{1}{\Lambda}\sum_{n=1}^\infty \int\limits_0^\infty kx^{k-1}c^{n*}(x)\,\mathrm{d}x =$$

$$= \frac{1}{\Lambda}\sum_{n=1}^\infty \int\limits_0^\infty kx^{k-1}\varrho^n\varsigma^{n*}(x)\,\mathrm{d}x = \frac{1}{\Lambda}\sum_{n=1}^\infty \varrho^n k\,\mathbb{E}\Big(\sum_{j=1}^n \vartheta_j\Big)^{k-1} \leqslant$$

$$\leqslant \frac{1}{\Lambda}\sum_{n=1}^\infty \varrho^n kn^{k-1}\frac{\Lambda\mathbb{E}\,\xi_i^k}{\varrho k} = \frac{\mathbb{E}\,\xi_i^k}{\varrho}\sum_{n=1}^\infty n^{k-1}\varrho^n = \frac{\mathbb{E}\,\xi_i^k}{\varrho} \times \varphi(\varrho, k-1).$$

Предложение 1 доказано. □

4. Оценки для процесса X_t

Вернёмся к исследуемому нами процессу X_t, описанному в разделе 2. Предполагаем, что в начальный момент времени $X_0 = (0,0)$, т. е. система свободна и начинается отсчёт времени до поступления первого требования. В момент поступления первого требования процесс X_t попадает в состояние $(1,0;0)$ — это момент регенерации.

Напомним, что описанный в разделе 3 процесс $\overset{\circ}{X}_t$ также стартует из этого состояния: $\overset{\circ}{X}_0 = (0,0)$. В момент поступления первого требования начинается первый период регенерации.

Определение 2. Случайная величина η не превосходит случайную величину θ *по распределению*, если при всех $s \in \mathbb{R}$ верно неравенство

$$F_\eta(s) = \mathsf{P}\{\eta \leqslant s\} \geqslant \mathsf{P}\{\theta \leqslant s\} = F_\theta(s),$$

откуда следует, что функция распределения случайной величины θ не превосходит функции распределения случайной величины η. Это отношение порядка, будем обозначать его $\eta \prec \theta$.

Утверждение 2. *$\eta \prec \theta$ тогда и только тогда, когда на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ можно построить случайные величины η' и θ', такие что $\eta' \leqslant \theta'$ при всех $\omega \in \Omega$ и $\mathsf{P}\{\eta \leqslant s\} = \mathsf{P}\{\eta' \leqslant s\}$ и $\mathsf{P}\{\theta \leqslant s\} = \mathsf{P}\{\theta' \leqslant s\}$ для всех $s \in \mathbb{R}$.*

Доказательство можно найти в [6].

Утверждение 3. *Если $X_0 = \overset{\circ}{X}_0 = (0,0)$ и выполнено условие* (7) (*напомним, что входящий поток для $\overset{\circ}{X}_t$ пуассоновский и имеет интенсивность Λ, а время*

обслуживания имеет функцию распределения (8)), *то при всех* $t \geqslant 0$ *выполнено соотношение*

$$n_t \prec \mathring{n}_t. \tag{14}$$

Доказательство. Для доказательства используется *метод одного вероятностного пространства*, который заключается в том, что на одном вероятностном пространстве строятся процессы X_t и $\mathring{X}_t$ (точнее говоря, их версии) таким образом, что на всех траекториях процессов выполнено неравенство $n_t \leqslant \mathring{n}_t$ при всех $t \geqslant 0$.

Эта конструкция возможна по двум причинам.

I. Входящий поток для процесса X_t имеет интенсивность, не превосходящую интенсивности входящего потока для $\mathring{X}_t$ ($\lambda(X_t) \leqslant \Lambda$ из условия (7)). Поэтому можно строить на одном вероятностном пространстве два входящих потока для обоих процессов: поток $\mathfrak{F}_t$ с интенсивностью $\lambda(X_t)$ для процесса X_t и поток $\mathfrak{F}_t^+$ с интенсивностью $\Lambda - \lambda(X_t)$; объединение потоков $\mathfrak{F}_t$ и $\mathfrak{F}_t^+$ — это поток с интенсивностью Λ для процесса $\mathring{X}_t$.

Таким образом, в каждый момент времени t число заявок, поступивших в СМО с процессом X_t не превосходит числа заявок, поступивших в СМО с процессом $\mathring{X}_t$. При этом в моменты поступления заявок в СМО с процессом X_t одновременно поступают заявки в СМО с процессом $\mathring{X}_t$; назовём такие заявки общими.

Иначе говоря, в каждый момент времени вектор $\left(x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\right)$ состоит из компонент вектора $\left(\mathring{x}_t^{(1)}, \mathring{x}_t^{(2)}, \ldots, \mathring{x}_t^{(\mathring{n}_t)}\right)$, т. е. $\left(x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\right)$ состоит из обслуженных времён находящихся на обслуживании требований, совпадающих с некоторыми из элементов вектора $\left(\mathring{x}_t^{(1)}, \mathring{x}_t^{(2)}, \ldots, \mathring{x}_t^{(\mathring{n}_t)}\right)$.

II. Интенсивность обслуживания i-го требования СМО, описываемой процессом $\mathring{X}_t = \left(\mathring{n}_t, \mathring{x}_t^{(0)}; \mathring{x}_t^{(1)}, \mathring{x}_t^{(2)}, \ldots, \mathring{x}_t^{(\mathring{n}_t)}\right)$, равна

$$\mathring{h}_i(\mathring{X}_t) = \frac{G'\left(\mathring{x}_t^{(i)}\right)}{1 - G\left(\mathring{x}_t^{(i)}\right)} = \frac{K}{1 + \mathring{x}_t^{(i)}}.$$

Таким образом, интенсивность обслуживания «общих» заявок у процесса $\mathring{X}_t$ не превосходит интенсивности обслуживания этих же заявок у процесса X_t, и это второе обстоятельство, позволяющее применить метод одного вероятностного пространства.

Учитывая, что время обслуживания имеет функцию распределения, которую можно восстановить по интенсивности обслуживания (см. (1), (2)), получаем, что остаточное время обслуживания «общих» заявок у процесса X_t не превосходит *по распределению* остаточного времени обслуживания этих заявок

у процесса $\overset{\circ}{X}_t$. Поэтому можно построить на одном вероятностном пространстве эти случайные величины так, что «общие» заявки завершат обслуживание у процесса X_t не позднее, чем у процесса $\overset{\circ}{X}_t$ (см. предложение 2).

Стало быть, в СМО с процессом $\overset{\circ}{X}_t$ поступает больший поток заявок, чем в СМО с процессом X_t, и одновременно поступившие (общие) заявки покидают СМО с процессом X_t раньше. Из этого следует утверждение предложения 3. $\square$

Следствие 1. *Из* (14) *следует, что свойство упорядоченности по распределению верно не только для n_t и $\overset{\circ}{n}_t$, но и для периодов занятости процессов X_t и $\overset{\circ}{X}_t$. Учитывая, что процессы X_t и $\overset{\circ}{X}_t$ регенерирующие (напомним, моментом регенерации является момент, когда процесс принимает значение* $(1,0;0)$, *т. е. момент начала периода занятости), можно сказать, что распределение периодов занятости ζ_i процесса X_t и периодов занятости $\overset{\circ}{\zeta}_i$ процесса $\overset{\circ}{X}_t$ не зависят от номера i. Таким образом, $\zeta_i \prec \overset{\circ}{\zeta}_t$ и для всех $k > 0$ верно неравенство*

$$\mathbb{E}(\zeta_i)^k \leqslant \mathbb{E}(\overset{\circ}{\zeta}_i)^k,$$

т. е. оценки (12) *верны и для ζ_i.*

Следствие 2. *Из замечания* 4 *и* (14) *следует, что при всех $t \geqslant 0$ (и $X_0 = (0,0)$) верно неравенство*

$$\mathsf{P}\{X_t \in \mathscr{S}_0\} = \mathsf{P}\{n_t = 0\} \geqslant \mathcal{P}_0(t) \geqslant e^{-\rho}. \tag{15}$$

5. Метод склеивания

Первоначально метод склеивания марковских процессов был применён к цепям Маркова с конечным числом состояний (см. [9, 19]). Основная идея этого метода состоит в следующем.

Пусть два однородных марковских процесса Y_t и Z_t с одинаковой переходной функцией в начальный момент времени имеют различные состояния; они заданы на $(\Omega, \mathscr{F}, \mathsf{P})$ и измеримы относительно фильтрации $\mathscr{F}_t$.

Обозначим

$$\tau = \tau(Y_0, Z_0) \stackrel{\text{def}}{=} \inf\{t > 0 \colon Y_t = Z_t\}.$$

Таким образом, $Y_\tau = Z_\tau$ и при $t > \tau$ распределения процессов Y_t и Z_t совпадают в соответствии с марковским свойством, т. е.

$$\mathsf{P}\{Y_t \in S \mid t > \tau\} = \mathsf{P}\{Z_t \in S \mid t > \tau\}$$

для всех $S \in \mathscr{B}(\mathscr{X})$, Y_t и Z_t, здесь $\mathscr{B}(\mathscr{X})$ — σ-алгебра на пространстве состояний $\mathscr{X}$ процессов Y_t и Z_t.

Имеет место *основное неравенство склеивания*:

$$\begin{aligned}|\mathsf{P}\{Y_t \in S\} - \mathsf{P}\{Z_t \in S\}| &= |\mathsf{P}\{Y_t \in S \,\&\, \tau > t\} - \mathsf{P}\{Z_t \in S \,\&\, \tau > t\}| + \\ &+ |\mathsf{P}\{Y_t \in S \,\&\, \tau < t\} - \mathsf{P}\{Z_t \in S \,\&\, \tau < t\}| = \\ &= |\mathsf{P}\{Y_t \in S \,\&\, \tau > t\} - \mathsf{P}\{Z_t \in S \,\&\, \tau > t\}| \leqslant \mathsf{P}\{\tau > t\}, \end{aligned} \tag{16}$$

поскольку при $\tau < t$ «склеивание» процессов Y_t и Z_t уже произошло, их распределения в момент t одинаковы и

$$\mathsf{P}\{Y_t \in S \mid \tau < t\} = \mathsf{P}\{Z_t \in S \mid \tau < t\};$$

здесь $S \in \mathscr{B}(\mathscr{X})$. Поэтому если для некоторой возрастающей положительной функции $\phi(t)$ удастся найти оценку $\mathbb{E}\,\phi\big(\tau(Y_0, Z_0)\big) \leqslant C(X_0, Y_0)$, то можно применить неравенство Маркова:

$$\mathsf{P}\{\tau(Y_0, Z_0) > t\} = \mathsf{P}\big\{\phi\big(\tau(Y_0, Z_0)\big) > \phi(t)\big\} \leqslant \frac{\mathbb{E}\,\phi\big(\tau(Y_0, Z_0)\big)}{\phi(t)} \leqslant \frac{C(X_0, Y_0)}{\phi(t)}. \tag{17}$$

Далее, если существует стационарное инвариантное распределение $\mathscr{P}$, к которому (слабо) сходятся распределения Y_t и Z_t при $t \to \infty$, то из (16) и (17) получаем

$$\left.\begin{aligned} &|\mathsf{P}\{Y_t \in S\} - \mathscr{P}(S)| \leqslant \int\limits_{\mathscr{X}} \frac{C(X_0, y)}{\phi(t)} \mathscr{P}(\,\mathrm{d}\,y) = \frac{\hat{C}(X_0)}{\phi(t)}, \\ &\text{следовательно,}\quad \|\mathscr{P}_t^{X_0} - \mathscr{P}\|_{TV} \stackrel{\text{def}}{=} \sup_{S \in \mathscr{B}(\mathscr{X})} |\mathscr{P}_t^{X_0}(S) - \mathscr{P}(S)| \leqslant 2\frac{\hat{C}(X_0)}{\phi(t)}. \end{aligned}\right\} \tag{18}$$

Изначально метод склеивания был предложен для цепей Маркова с дискретным множеством состояний [9], однако изучаемый нами процесс X_t с переходными вероятностями (6) — это процесс в непрерывном времени для двух независимых процессов X_t и X'_t, удовлетворяющих (6), но с различными начальными состояниями $\mathsf{P}\{\tau(X_0, x'_0) < \infty\} = 0$. Поэтому мы будем использовать понятие *успешной склейки*, предложенное в [13].

Определение 3. Пусть X_t и X'_t — две независимые версии одного марковского процесса с пространством состояний $\mathscr{X}$: X_t с начальным состоянием X_0, X'_t с начальным состоянием X'_0.

Успешной склейкой (см. [13]) процессов X_t и X'_t называется парный процесс $\mathcal{Z}_t = (Y_t, Y'_t)$, заданный на некотором вероятностном пространстве, такой что

1) для всех $s \geqslant 0$ и $S \in \mathscr{B}(\mathscr{X})$ выполнены равенства

$$\mathbf{P}\{Y_t \in S\} = \mathbf{P}\{X_t \in S\}, \quad \mathbf{P}\{Y'_t \in S\} = \mathbf{P}\{X'_t \in S\}$$

(соответственно $Y_0 = X_0$ и $Y'_0 = X'_0$). При этом для любого фиксированного момента $t \geqslant 0$ случайные величины Y_t, Y'_t, X_t и X'_t рассматриваются как некоторые случайные величины, но не как случайные процессы! Таким образом, речь идёт не о совпадении распределений двух процессов, под которым понимается совпадение всех конечномерных распределений;

2) для всех $t > \tau = \tau(X_0, X_0') = \tau(Y_0, Y_0') \stackrel{\text{def}}{=} \inf\{t \geqslant 0\colon Y_t = Y_t'\}$ выполняется равенство $Y_t = Y_t'$;
3) для любых начальных значений $X_0, X_0' \in \mathscr{X}$ верно $\mathbf{P}\{\tau(X_0, X_0') < \infty\} = 1$.

Для успешной склейки можно также использовать основное неравенство склеивания (16):

$$|\mathsf{P}\{X_t \in S\} - \mathsf{P}\{X_t' \in S\}| = |\mathsf{P}\{Y_t \in S\} - \mathsf{P}\{Y_t' \in S\}| = \\ = |\mathsf{P}\{Y_t \in S\} - \mathsf{P}\{Y_t' \in S\}| \times \big(\mathbf{1}(\tau > t) + \mathbf{1}(\tau \leqslant t)\big) \leqslant \mathsf{P}\{\tau > t\}.$$

Итак, наша цель — сконструировать успешную склейку для двух версий процесса X_t (X_t с начальным состоянием $X_0 = (0,0)$ и X_t' с начальным состоянием $X_0' = \big(n_0', x_t'^{(0)}; x_t'^{(1)}, x_t'^{(2)}, \ldots, x_t'^{(n_0')}\big)$), причём каждый из процессов удовлетворяет условиям (6), и из этого получить оценку скорости сходимости распределения процесса X_t к стационарному распределению.

Для применения метода склеивания нам понадобится лемма.

5.1. Основная лемма склеивания

Этот факт изложен во многих публикациях (см., например, [16,23]), но здесь мы приведём его простейшую форму.

Определение 4. Общей частью распределения двух случайных величин ξ_1 и ξ_2 с функциями распределения $\Psi_j(s) = \mathbf{P}\{\xi_j \leqslant s\}$, $s \in \mathbb{R}$, $j = 1, 2$, назовём

$$\varkappa \stackrel{\text{def}}{=} \varkappa\big(\Psi_1(s), \Psi_2(s)\big) \stackrel{\text{def}}{=} \int\limits_{-\infty}^{\infty} \min\big(\psi_1(u), \psi_2(u)\big)\,\mathrm{d}u,$$

где

$$\psi_j(s) \stackrel{\text{def}}{=} \begin{cases} \Psi_j'(s), & \text{если существует } \Psi_j'(s), \\ 0 & \text{в противном случае;} \end{cases}$$

здесь $j = 1, 2$.

Лемма 1. *Если $\varkappa > 0$, то* на некотором вероятностном пространстве *можно построить две случайные величины ξ_1' и ξ_2', такие что*

$$\mathsf{P}\{\xi_j' \leqslant s\} = \mathsf{P}\{\xi_j \leqslant s\} = \Psi_j; \tag{19}$$

$$\mathsf{P}\{\xi_1' = \xi_2'\} \geqslant \varkappa. \tag{20}$$

Доказательство. Обозначим

$$\psi(u) \stackrel{\text{def}}{=} \min\big(\psi_1(u), \psi_2(u)\big), \quad \Psi(s) \stackrel{\text{def}}{=} \int\limits_{-\infty}^{s} \psi(u)\,\mathrm{d}u,$$

$$\hat{\Psi}(s) \stackrel{\text{def}}{=} \frac{\Psi(s)}{\varkappa}, \quad \hat{\Psi}_j(s) \stackrel{\text{def}}{=} \frac{\Psi_j(s) - \Psi(s)}{1 - \varkappa}, \quad j = 1, 2$$

(если $\varkappa = 1$, то $\hat{\Psi}_j(s) \equiv 0$).

Пусть $\mathscr{U}_1$, $\mathscr{U}_2$ и $\mathscr{U}_3$ — независимые случайные величины, равномерно распределённые на отрезке $[0;1]$. Положим

$$\xi'_j \stackrel{\text{def}}{=} \hat{\Psi}^{-1}(\mathscr{U}_j) \times \mathbf{1}(\mathscr{U}_3 \leqslant \varkappa) + \hat{\Psi}_j^{-1}(\mathscr{U}_j) \times \mathbf{1}(\mathscr{U}_3 > \varkappa).$$

Напомним, что под обратной функцией монотонной функции $f(s)$ мы понимаем

$$f^{-1}(u) \stackrel{\text{def}}{=} \inf\{s\colon f(s) \geqslant u\}.$$

Случайные величины ξ'_j удовлетворяют условиям (19), (20).

Лемма доказана. □

Замечание 5. Если $n_\vartheta = 0$ в некоторый момент времени ϑ (СМО свободна), то остаточная длина свободного периода процесса X_t (описанного в разделе 2) имеет плотность распределения

$$\varphi(s) = \lambda(X_{\vartheta+s}) \exp\biggl(\int\limits_0^s \lambda(X_{\vartheta+u})\,\mathrm{d}u\biggr) \in (\lambda_0 e^{-\Lambda s}; \Lambda_0 e^{-\lambda s}).$$

Поэтому если рассматриваются два процесса X_t и X'_t с различными начальными состояниями X_0 и X'_0 и в некоторый момент времени ϑ они оба оказались в множестве $\mathscr{S}_0$ (СМО свободны), то общая часть $\varkappa_\vartheta$ распределений их остаточных времён пребывания в множестве $\mathscr{S}_0$ удовлетворяет неравенству

$$\hat{\varkappa}_\vartheta \geqslant \int\limits_0^\infty \lambda_0 e^{-\Lambda s}\,\mathrm{d}s = \frac{\lambda_0}{\Lambda} \stackrel{\text{def}}{=} \hat{\varkappa}. \tag{21}$$

6. Конструирование успешной склейки для пары процессов X_t, X'_t

Рассмотрим два независимых процесса: процесс X_t с начальным состоянием $X_0 = (0,0)$ и процесс X'_t с начальным состоянием

$$X'_0 = \bigl(n'_0, x'^{(0)}_0; x'^{(1)}_0, x'^{(2)}_0, \ldots, x'^{(n'_0)}_0\bigr).$$

Будем строить их склейку $\mathscr{Z}_t = (Y_t, Y'_t)$ на вероятностном пространстве

$$(\Omega, \mathscr{F}, \mathsf{P}) = \prod_{i=0}^{\infty} (\Omega_i, \mathscr{F}_i, \mathsf{P}_i),$$

где $\Omega_i = [0;1]$, $\mathscr{F}_i = \mathscr{B}([0;1])$ и P_i — мера Лебега на $[0;1]$.

Опишем процедуру конструирования склейки по шагам, которые производятся в моменты изменения первой компоненты любого из процессов $Y_t = \bigl(m_t, y_t^{(0)}; y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(m_t)}\bigr)$, $Y'_t = \bigl(m'_t, y'^{(0)}_t; y'^{(1)}_t, y'^{(2)}_t, \ldots, y'^{(m'_t)}_t\bigr)$; обозначим эти моменты t_k:

$$t_0 \stackrel{\text{def}}{=} 0; \quad t_k \stackrel{\text{def}}{=} \inf\{t > t_{k-1}\colon m_t + m'_t \neq m_{t_{k-1}} + m'_{t_{k-1}}\}.$$

На каждом шаге строятся остаточные *виртуальные* времена обслуживания находящихся в СМО заявок (в предположении, что до окончания обслуживания данной заявки первые компоненты — n_t и n'_t — обоих процессов не изменятся) и остаточные *виртуальные* времена до прихода следующей заявки. Для этого используется некоторое количество ν_k экземпляров $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$, так что на каждом шаге используются новые $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$.

Опишем алгоритм конструирования.

А. Если $m_{t_k} + m'_{t_k} > 0$, для удобства положим

$$Y_{\vartheta_k} = (\imath, \alpha_0; \alpha_1, \alpha_2, \ldots, \alpha_\imath), \quad Y'_{\vartheta_k} = (\jmath, \beta_0; \beta_1, \beta_2, \ldots, \beta_\jmath).$$

На k-м шаге строятся случайные величины на пространствах $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$, $i = \ell_k + 1, \ell_k + 2, \ldots, \ell_{k+1}$; при этом

$$\nu_k \stackrel{\text{def}}{=} \ell_{k+1} - \ell_k = n_{\vartheta_k} + n'_{\vartheta_k} + 2 = \imath + \jmath + 2.$$

К этому моменту времени было использовано

$$\ell_k = \sum_{i=0}^{k-1} \nu_i$$

экземпляров пространства $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$.

В момент t_k *виртуальное* время до поступления очередного требования ξ (при условии, что не закончится обслуживание какого-то требования) для Y_t имеет функцию распределения

$$F_1(s) = 1 - \exp\biggl(\int_0^s \lambda(\imath, \alpha_0 + u; \alpha_1 + u, \alpha_2 + u, \ldots, \alpha_\imath + u)\,\mathrm{d}u\biggr).$$

Строим на пространстве $(\Omega_{\ell_k+1}, \mathscr{F}_{\ell_k+1}, \mathsf{P}_{\ell_k+1})$ случайную величину

$$\xi \stackrel{\text{def}}{=} F_1^{-1}(\omega), \quad \omega \in \Omega_{\ell_k+1}.$$

В момент t_k *виртуальное* время до окончания обслуживания j-го ($j = 1, 2, \ldots, \imath$) требования η_j (при условии, что не закончится обслуживание какого-то другого требования и не поступит новых требований) Y_t имеет функцию распределения

$$G_{1,j}(s) = 1 - \exp\biggl(\int_0^s h_j(\ell, \alpha_0 + u; \alpha_1 + u, \alpha_2 + u, \ldots, \alpha_\imath + u)\,\mathrm{d}u\biggr)$$

Строим на пространстве $(\Omega_{\ell_k+j+1}, \mathscr{F}_{\ell_k+j+1}, \mathsf{P}_{\ell_k+j+1})$ случайные величины

$$\eta_j \stackrel{\text{def}}{=} G_{1,j}^{-1}(\omega), \quad \omega \in \Omega_{\ell_k+j+1}.$$

Аналогично строятся случайные величины ξ' и η'_i $(i = 1, 2, \ldots, \jmath)$:

$$F_2(s) = 1 - \exp\biggl(\int\limits_0^s \lambda(\jmath, \beta_0 + u; \beta_1 + u, \beta_2 + u, \ldots, \beta_\jmath + u)\,\mathrm{d}u\biggr),$$

$$\xi' \stackrel{\text{def}}{=} F_2^{-1}(\omega), \quad \omega \in \Omega_{\ell_k + \imath + 1};$$

$$G_{2,j}(s) = 1 - \exp\biggl(\int\limits_0^s h_j(\jmath, \beta_0 + u; \beta_1 + u, \beta_2 + u, \ldots, \beta_\jmath + u)\,\mathrm{d}u\biggr),$$

$$\eta'_j \stackrel{\text{def}}{=} G_{2,j}^{-1}(\omega), \quad \omega \in \Omega_{\ell_k + \imath + j + 2}.$$

Использовано $\nu_k = \imath + \jmath + 2$ экземпляров $(\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$; $\ell_{k+1} = \ell_k + \nu_k$.

В момент времени

$$t_{k+1} = t_k + \min\{\xi, \xi', \eta_j, \eta'_i \ (j = 1, 2, \ldots, \imath,\ i = 1, 2, \ldots, \jmath)\}$$

произойдёт изменение первой компоненты одного из процессов Y_t и Y'_t.

Если опять $m_{t_k} + m'_{t_k} > 0$, повторяем процедуру А.

Б. Если $m_{t_k} + m'_{t_k} = 0$, то $Y_{t_k} = (0, \alpha_0)$, $Y'_{t_k} = (0, \beta_0)$. Функции распределения времён ξ и ξ' до прихода следующей заявки для процессов Y_t и Y'_t соответственно имеют вид

$$\mathsf{P}\{\xi \leqslant s\} = F_1(s) = 1 - \exp\biggl(\int\limits_0^s \lambda(0, \alpha_0 + u)\,\mathrm{d}u\biggr);$$

$$\mathsf{P}\{\xi' \leqslant s\} = F_2(s) = 1 - \exp\biggl(\int\limits_0^s \lambda(0, \beta_0 + u)\,\mathrm{d}u\biggr).$$

В соответствии с замечанием 5 и (21) общая часть этих распределений не меньше

$$\hat{\varkappa} = \frac{\lambda_0}{\Lambda}.$$

Поэтому на пространстве

$$\prod_{i=\ell_k+1}^{\ell_k+3} (\Omega_i, \mathscr{F}_i, \mathsf{P}_i)$$

(здесь $\nu_k = 3$) в соответствии с леммой 1 строим ξ и ξ' таким образом, что $\mathsf{P}\{\xi = \xi'\} \geqslant \hat{\varkappa}$.

В момент времени $t_{k+1} \stackrel{\text{def}}{=} t_k + \min\{\xi, \xi'\}$ первая компонента хотя бы одного из процессов Y_t, Y'_t возрастёт, и с вероятностью большей чем $\hat{\varkappa}$ первые компоненты процессов Y_t, Y'_t возрастут одновременно, т. е. произойдёт событие $Y_{t_{k+1}} = Y'_{t_{k+1}} = (1, 0; 0)$.

B. Если случилось событие $Y_{t_k} = Y'_{t_k} = (1,0;0)$, при $t > t_k$ полагаем процессы Y_t и Y'_t равными и строим процесс Y_t (и равный ему Y'_t) по процедуре, описанной в пункте A.

Несложно заметить, что сконструированные по алгоритму A—B процессы Y_t и Y'_t удовлетворяют условиям 1), 2) определения 3.

Покажем, что выполнено также условие 3) определения 3.

Рассмотрим процесс $Y'_t \overset{\mathscr{D}}{=} X'_t$. Если $n'_0 > 0$, то, пользуясь теми же соображениями, что и в разделах 3 и 4, можно показать, что остаточное время ζ_1 первого периода занятости процесса Y'_t, зависящее от начального состояния Y'_0, имеет $k \in [0;K)$ конечных моментов.

Обозначим

$$\vartheta_0 \overset{\text{def}}{=} \zeta_0, \quad \vartheta_k \overset{\text{def}}{=} \inf\{t > \vartheta_k \colon Y'_t \in \mathscr{S}_0,\ Y'_{t-0} \notin \mathscr{S}_0\};$$

ϑ_k — момент окончания k-го периода занятости процесса Y'_t. В случае $Y'_0 \in \mathscr{S}$ положим $\vartheta_0 \overset{\text{def}}{=} 0$.

Поскольку распределение процесса Y_t совпадает с распределением процесса X_t, верно неравенство

$$\mathsf{P}\{X_{\vartheta_k} \in \mathscr{S}_0\} \geqslant e^{-\rho}$$

(см. (15)).

Как мы уже видели, после совместного попадания процессов Y_t и Y'_t в множество $\mathscr{S}$ с вероятностью большей чем $\hat{\varkappa}$ при выходе из этого множества процессы «склеиваются». Поэтому, снова обозначив

$$\tau \overset{\text{def}}{=} \inf\{t > 0 \colon Y_t = Y'_t\},$$

η_k — длина свободного периода Y'_t после ϑ_k, получаем, что

$$\mathsf{P}\{\tau > \vartheta_k + \eta_k\} \leqslant (1 - \varpi)^k.$$

Заметим, что $\tau = \tau(Y'_0) = \tau(X'_0)$ (поскольку $X_0 = Y_0 = (0,0)$ зафиксировано).

Очевидно,

$$\vartheta_k + \eta_k = \vartheta_0 + \eta_0 + \sum_{i=1}^{k} (\zeta_i + \eta_i),$$

где ζ_i — i-й период занятости, а η_i — следующий за ним свободный период.

Обозначим

$$\mathscr{E}_k \overset{\text{def}}{=} \{Y_{\vartheta_k} \neq Y'_{\vartheta_k}\ \&\ Y_{\vartheta_k+\eta_k} = Y'_{\vartheta_k+\eta_k}\};$$

это событие, заключающееся в том, что склеивание произошло сразу после k-го попадания Y'_t в множество $\mathscr{S}_0$.

Утверждение 4. *При $r \in [1;K)$ выполнено*

$$\mathbb{E}\,\tau^r \leqslant C_1(\varpi, r)\mathbb{E}\,\vartheta_0^r + C_1(\varpi, r)\mathbb{E}\,\zeta_1^r + C_3(\varpi, r, \lambda_0, \Lambda),$$

константы C_1, C_2, C_3 вычисляются по приведённым ниже формулам.

Доказательство. Сначала оценим $\mathbb{E}\,(\eta_k)^r|_{\mathscr{E}_k}$. В случае события $\mathscr{E}_k$ случайная величина η_k (в соответствии с леммой 1) имеет плотность распределения

$$\varphi(s) = \min\Bigg\{\lambda(Y_{\vartheta_k}+s)\exp\Bigg(-\int\limits_0^s \lambda(Y_{\vartheta_k}+u)\,\mathrm{d}\,u\Bigg),$$
$$\lambda(Y'_{\vartheta_k}+s)\exp\Bigg(-\int\limits_0^s \lambda(Y'_{\vartheta_k}+u)\Bigg)\Bigg\} \leqslant \Lambda e^{-\lambda_0 s}.$$

Поэтому

$$\mathbb{E}\,(\eta_k)^r|_{\mathscr{E}_k} \leqslant \frac{r!\,\Lambda}{\lambda_0^{r+1}}.$$

Легко заметить, что и

$$\mathbb{E}\,(\eta_k)^r|_{\bar{\mathscr{E}}_k} \leqslant \frac{r!\,\Lambda}{\lambda_0^{r+1}};$$

ζ_k не зависят от $\mathscr{E}_j$ при $k \neq j$.

Оценим $\mathbb{E}\,\tau^r$ (используя неравенство Йенсена в форме (13)):

$$\mathbb{E}\,\tau^r \leqslant \sum_{k=1}^{\infty}\mathbb{E}\Bigg(\vartheta_0+\eta_0+\sum_{i=1}^{k}(\zeta_i+\eta_i)\Bigg)^r \mathbf{1}(\mathscr{E}_k) \leqslant$$
$$\leqslant \sum_{k=1}^{\infty}(1-\varpi)^{k-1}\varpi(2k+2)^{r-1}\left(\left(\mathbb{E}\,\vartheta_0^r+(k+1)\frac{r!\,\Lambda}{\lambda_0^{r+1}}+k\mathbb{E}\,\zeta_1^r\right)\right) <$$
$$< 2^{r-1}\Bigg(\mathbb{E}\,\vartheta_0^r\sum_{k=1}^{\infty}(1-\varpi)^{k-1}(k+1)^{r-1}+\frac{r!\,\Lambda}{\lambda_0^{r+1}}\sum_{k=1}^{\infty}(1-\varpi)^{k-1}(k+1)^r+$$
$$+\mathbb{E}\,\zeta_1^r\sum_{k=1}^{\infty}(1-\varpi)^{k-1}(k+1)^{r-1}k\Bigg).$$

Предложение 4 доказано. □

Итак, сконструированная выше пара $\mathscr{Z}_t = (Y_t, Y'_t)$ — успешная склейка процессов X_t и X'_t, и получена явная оценка времени склеивания:

$$\mathbb{E}\left(\tau(X'_0)\right)^r \leqslant \mathcal{C}(r,\lambda_0,\Lambda,\mathbb{E}\,\vartheta_0^r). \tag{22}$$

т. е.

$$|\mathsf{P}\{X_t\in S\}-\mathsf{P}\{X'_t\in S\}| \leqslant \frac{\mathcal{C}(r,\lambda_0,\Lambda,\mathbb{E}\,\vartheta_0^r)}{t^r}, \tag{23}$$

однако следующий шаг — интегрирование (18) — мы произвести не можем, поскольку стационарное распределение СМО $M|G|\infty$ (процесс $\overset{\circ}{X}_t$), и тем более $G|G|\infty$ (процесс X_t), *неизвестно*.

7. Процесс X_t как регенерирующий процесс

Напомним определение регенерирующего процесса.

Определение 5. Случайный процесс $\{W_t,\ t \geqslant 0\}$, заданный на вероятностном пространстве $(\Omega, \mathscr{F}, \mathbf{P})$, измеримый относительно фильтрации $\{\mathscr{F}_t,\ t \geqslant 0\}$ и принимающий значения в измеримом пространстве состояний $\big(\mathscr{W}, \mathscr{B}(\mathscr{W})\big)$ называется *регенерирующим процессом*, если существует последовательность марковских по отношению к фильтрации $\mathscr{F}_t$ моментов $\{\theta_i\}_{i\in\mathbb{N}}$, такая что

1) $W_{\theta_i} = W_{\theta_j}$ при всех $i, j \in \mathbb{N}$;
2) случайные элементы $\Xi_i \stackrel{\text{def}}{=} \{W_t,\ t \in [\theta_i, \theta_{i+1}]\}$ $(i \in \mathbb{N})$ одинаково распределены и независимы в совокупности.

Назовём периодами регенерации промежутки $[\theta_i, \theta_{i+1}]$; очевидно, длины периодов регенерации $\chi_i \stackrel{\text{def}}{=} \theta_{i+1} - \theta_i$ одинаково распределены и независимы в совокупности. Обозначим $\Phi(s) \stackrel{\text{def}}{=} \mathsf{P}\{\chi_i \leqslant s\}$ (неизвестную нам) функцию распределения случайной величины χ_i.

Известно, что если $\mathbb{E}\,\chi_i < \infty$, то распределение регенерирующего процесса W_t слабо сходится к стационарному (инвариантному) распределению.

Заметим, что для введённого в разделе 2 процесса

$$X_t = \big(n_t, x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\big)$$

периоды регенерации разделяются на период занятости и свободный период: $\chi_k = \zeta_k + \eta_k$.

Моменты периода занятости оцениваются с помощью (11) и (12).

Моменты свободных периодов η_i, следующих за периодами занятости ζ_i, оценены в предложении 4. Кроме того, период регенерации заведомо больше своей части — свободного периода, а плотность распределения свободного периода $\varphi(s)$ не меньше $\lambda_0 e^{-\Lambda s}$ (замечание 5).

Таким образом, известны оценки

$$\mathbb{E}\,\chi_k^r \leqslant 2^{r-1}(\mathbb{E}\,\zeta_1^r + \mathbb{E}\,\eta_1^r) = \mathcal{M}_r, \quad \mathbb{E}\,\chi_i > \frac{\lambda_0}{\Lambda^2}. \tag{24}$$

Марковские моменты $\{\theta_i\}_{i\in\mathbb{N}}$ образуют вложенный процесс восстановления $N_t = \sum\limits_{i=1}^{\infty} \mathbf{1}(\theta_i < t)$.

Рассмотрим *перескок* процесса N_t, а именно марковский процесс $B_t \stackrel{\text{def}}{=} t - \theta_{N_t}$ (с пространством состояний $\mathbb{R}_+$), распределение которого $\mathfrak{P}_t$, как известно [2, 3], слабо сходится к стационарному инвариантному распределению $\mathfrak{P}$ с функцией распределения

$$\tilde{\Phi}(s) = \frac{\int\limits_0^s \big(1 - \Phi(u)\big)\,\mathrm{d}\,u}{\mathbb{E}\,\chi_i}. \tag{25}$$

К этому же распределению (25) слабо сходится распределение *недоскока* $R_t \stackrel{\text{def}}{=} \theta_{N_t+1} - t$ процесса восстановления N_t.

Поскольку процесс X_t регенерирующий, его распределение полностью определяется временем, прошедшим с момента начала периода регенерации, т. е. $\mathscr{P}_t = \mathbb{F}(B_t)$.

Теперь рассмотрим процесс X_t с начальным состоянием $(0,0)$ и процесс X'_t с произвольным начальным состоянием X_0. Этим процессам соответствуют вложенные процессы восстановления N_t и N'_t, а процессам N_t и N'_t соответствуют процессы перескоков B_t и B'_t и процессы недоскоков R_t и R'_t.

При конструировании успешной склейки $\mathscr{Z}_t = (Y_t, Y'_t)$ для процессов X_t и X'_t одновременно конструируется успешная склейка $\mathscr{W}_t = (D_t, D'_t)$ процессов B_t и B'_t. Обозначим через $\mathfrak{P}_t$ и $\mathfrak{P}'_t$ распределение процессов B_t и B'_t соответственно ($\mathsf{P}\{B_t \in S\} = \mathfrak{P}_t(S)$ и $\mathsf{P}\{B'_t \in S\} = \mathfrak{P}'_t(S)$).

Из (22) и (23) сразу следует, что

$$|\mathsf{P}\{B_t \in S\} - \mathsf{P}\{B'_t \in S\}| \leqslant \frac{\mathcal{C}(\varpi, r, \lambda_0, \Lambda, \mathbb{E}\,\vartheta_0^r)}{t^r}, \tag{26}$$

где $\vartheta_0 = R'_0$ — недоскок процесса N'_t в момент времени $t = 0$.

Утверждение 5.

$$\|\mathfrak{P}_t - \mathfrak{P}\|_{TV} \leqslant \\ \leqslant \frac{2\left(C_1(\varpi, r)\frac{\mathbb{E}\,\chi_i^{r+1}}{(r+1)\mathbb{E}\,\chi_i} + C_1(\varpi, r)\mathbb{E}\,\zeta_1^r + C_3(\varpi, r, \lambda_0, \Lambda)\right)}{t^r} = \frac{\mathcal{C}_1(r, \lambda_0, \Lambda)}{t^r},$$

где $r \in [1, K-1)$.

Доказательство. Рассмотрим процесс перескока B_t с начальным условием $B_0 = 0$ и процесс перескока $\tilde{B}_t$, имеющий стационарное начальное распределение $\mathfrak{P}$; при всех $t \geqslant 0$

$$\mathsf{P}\{\tilde{B}_t \in S\} = \mathfrak{P}(S) = \int\limits_S \frac{1 - \Phi(u)\,\mathrm{d}u}{\mathbb{E}\,\chi_i}. \tag{27}$$

Подставляя в (26) вместо $\vartheta_0 = R'_0$ стационарный недоскок $\tilde{R}_0$ с распределением (25), получаем утверждение предложения 5.

Учитывая (24), можно вычислить верхнюю оценку величины $\mathcal{C}_1(r, \lambda_0, \Lambda)$. □

Теорема 2. *В условиях* (7) *верно неравенство*

$$\|\mathscr{P}_t - \mathscr{P}\|_{TV} \leqslant \frac{\mathcal{C}_1(r, \lambda_0, \Lambda)}{t^r}$$

где $r \in [1, K-1)$.

Доказательство. Как уже говорилось, $\mathscr{P}_t = \mathbb{F}(\mathfrak{P}_t)$. Соответственно, $\mathscr{P} = \mathbb{F}(\mathfrak{P})$. Напомним, что χ_i — длина периода регенерации, включающего в себя абсолютно непрерывные величины, т. е. распределение случайной величины χ_i

имеет плотность. Поэтому имеет плотность $\psi_t(s)$ перескок B_t; как уже говорилось (см. (27)), известна функция распределения стационарного перескока, и её плотность равна

$$\psi(s) = \frac{1 - \Phi(u)}{\mathbb{E}\,\chi_i}.$$

Обозначим $\sigma_t = \{s\colon \psi_t(s) > \psi(s)\}$.

Оценим

$$\begin{aligned}
&|\mathscr{P}_t(S) - \mathscr{P}(S)| = \\
&\quad = \left| \int\limits_0^\infty \mathsf{P}\{X_t \in S \mid B_t = u\}\psi_t(u)\,\mathrm{d}u - \int\limits_0^\infty \mathsf{P}\{X_t \in S \mid B_t = u\}\psi(u)\,\mathrm{d}u \right| \leqslant \\
&\quad \leqslant \max\left\{ \int\limits_{\sigma_t} \big(\psi_t(u) - \psi(u)\big)\,\mathrm{d}u, \int\limits_{\mathbb{R}_+ \setminus \sigma_t} \big(\psi(u) - \psi_t(u)\big)\,\mathrm{d}u \right\} = \\
&\quad = \max\{|\mathfrak{P}_t(\sigma_t) - \mathfrak{P}(\sigma_t)|, |\mathfrak{P}_t(\mathbb{R}_+ \setminus \sigma_t) - \mathfrak{P}(\mathbb{R}_+ \setminus \sigma_t)|\} \leqslant \\
&\quad \leqslant \sup_{\sigma \in \mathscr{B}(\mathbb{R})} |\mathfrak{P}_t(\sigma) - \mathfrak{P}(\sigma)|.
\end{aligned}$$

Тогда $\|\mathscr{P}_t - \mathscr{P}\|_{TV} \leqslant \|\mathfrak{P}_t - \mathfrak{P}\|_{TV}$, что завершает доказательство. □

Замечание 6. Полученная оценка неоптимальна по следующим причинам.

1. Применение метода склеивания для процессов в непрерывном времени, как правило, даёт неоптимальную оценку.
2. При конструировании успешной склейки использовались только моменты попадания процесса X'_t в свободное состояние; также можно было бы учесть и моменты попадания в свободное состояние процесса X_t.
3. Условия (7) не позволяют аккуратно оценить получаемые в ходе вычислений величины ($\mathcal{C}$ и $\mathcal{C}_1$). При исследовании конкретных прикладных задач эти величины могут быть оценены точнее.

Тем не менее полученная оценка может быть полезна в практических приложениях, поскольку имитационное моделирование не всегда способно дать достаточно качественные оценки.

Автор благодарит Л. Г. Афанасьеву, А. Ю. Веретенникова и А. Д. Маниту за важные и полезные замечания при обсуждении представленной работы. Также автор благодарит Э. Ю. Калимулину за большую помощь в подготовке текста. Работа выполнена при финансовой поддержке РФФИ (проект № 17-01-00633 А).

Литература

[1] Боровков А. А. Асимптотические методы в теории массового обслуживания. — М.: Физматгиз, 1980.

[2] Гнеденко Б. В., Беляев Ю. К., Соловьёв А. Д. Математические методы в теории надёжности. — М.: Наука, 1965.

[3] Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. — М.: Наука, 1966.

[4] Севастьянов Б. А. Эргодическая теорема для марковских процессов и её приложение к телефонным системам с отказами. — Теор. вер. и её прим. — 1957. — Т. 2, № 1. — С. 106—116.

[5] Севастьянов Б. А. Формулы Эрланга в телефонии при произвольном законе распределения длительности разговора // Теор. вер. и её прим. — 1957. — Т. 2, № 1. — С. 135.

[6] Штойян Д. Качественные свойства и оценки стохастических моделей. — М.: Мир, 1979.

[7] Afanasyeva L. G., Tkachenko A. V. On the convergence rate for queueing and reliability models described by regenerative processes // J. Math. Sci. — 2016. — Vol. 218, no. 2. — P. 119—136.

[8] Asmussen S. Applied Probability and Queues. — New York: Springer, 2003.

[9] Doeblin W. Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états // Rev. Math. l'Union Interbalkanique. — 1938 — Vol. 2. — P. 77—105.

[10] Erlang A. K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges // P. O. Elect. Eng. J. — 1918. — Vol. 10. — P. 189—197.

[11] Fortet R. Calcul des probabilités. — Paris: CNRS, 1950.

[12] Golovastova E. On convergence rate for an infinite-channel queuing system with Poisson input flow. — `arXiv:1711.07075[math.PR]`.

[13] Griffeath D. A maximal coupling for Markov chains // Z. Wahrsch. Verw. Gebiete. — 1975 — Vol. 31, no. 2. — P. 95—106.

[14] Kalashnikov V. V. Mathematical Methods in Queueing Theory. — Dordrecht: Kluwer Academic, 1994.

[15] Karlin S., Taylor H. M. A Second Course in Stochastic Processes. — San Diego: Academic Press, 1975.

[16] Kato K. Coupling lemma and its application to the security analysis of quantum key distribution // Tamagawa Univ. Quantum ICT Research Inst. Bull. — 2014. — Vol. 4, no. 1. — P. 23—30.

[17] Stadje W. The busy period of the queueing system $M|G|\infty$ // J. Appl. Probab. — 1985 — Vol. 22. — P. 697—704.

[18] Takács L. Introduction to the Theory of Queues. — Oxford: Oxford Univ. Press, 1962.

[19] Thorisson H. Coupling, Stationarity, and Regeneration. — Berlin: Springer, 2000.

[20] Veretennikov A. Yu. On the rate of convergence for infinite server Erlang–Sevastyanov's problem // Queueing Systems. — 2014. — Vol. 76, no. 2. — P. 181—203.

[21] Veretennikov A. On recurrence and availability factor for single–server system with general arrivals // Reliability Theory Appl. — 2016. — Vol. 11, no. 3 (42). — P. 49—58.

[22] Veretennikov A. On convergence rate for Erlang—Sevastyanov type models with infinitely many servers // Theory Stoch. Processes. — 2017. — No. 1. — P. 88—102. — `https://arxiv.org/abs/1412.3849`.

[23] Veretennikov A., Butkovsky O. A. On asymptotics for Vaserstein coupling of Markov chains // Stoch. Processes Appl. — 2013. — Vol. 123, no. 9. — P. 3518—3541.

[24] Veretennikov A. Yu., Zverkina G. A. Simple proof of Dynkin's formula for single-server systems and polynomial convergence rates // Markov Processes Relat. Fields. — 2014. — Vol. 20. — P. 479—504.