НЕРАВЕНСТВО ЛОРДЕНА И СКОРОСТЬ СХОДИМОСТИ РАСПРЕДЕЛЕНИЯ ОДНОЙ ОБОБЩЁННОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ ЭРЛАНГА – СЕВАСТЬЯНОВА[1]

Зверкина Г. А.[2]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

Во многих прикладных задачах теории надежности и массового обслуживания очень важно не только доказывать существование стационарного распределения, но и уметь оценивать скорость сходимости распределения к стационарному. Стандартные методы получения таких оценок предполагают, что времена обслуживания (ремонта или работы) экспонециальны, входящий поток – пуассоновский и все формирующие процесс обслуживания (надёжности) случайные величины (сл.в.) независимы. Результаты для таких простейших случаев хорошо известны. Отказ от предположений независимости и экспоненциальности этих сл.в. приводит к довольно сложным случайным процессам, которые очень трудно изучать с помощью стандартных процедур. Для таких процессов нужно использовать более сложную технику. Для этого потребуется некоторое обобщение (и доказательство) ряда известных результатов. Один из таких результатов – обобщённое неравенство Лордена, используемое в данной статье. «Классическое» неравенство Лордена касается «классических» процессов восстановления. В работе используется обобщение этого неравенства для случая «слабо зависимых» и имеющих в некотором смысле «близкие» распределения интервалов между моментами восстановления. Такое обобщение позволяет изучать скорость сходимости для широкого класса сложных процессов в ТМО и в смежных дисциплинах. В данной работе изучается одна обобщённая система массового обслуживания Эрланга – Севастьянова.

Ключевые слова: регенерирующие марковские процессы, метод склеивания, метрика полной вариации, обобщённая система Эрланга – Севастьянова, скорость сходимости.

---

[1] *Работа выполнена при финансовой поддержке РФФИ, грант №20-01-00575А.*
Автор признателен Э.Ю. Калимулиной за ценное обсуждение содержания статьи.

[2] *Галина Александровна Зверкина, к.ф.-м.н., доцент* (*zverkina@gmail.com*).

1. Введение

Формулы Эрланга дают явные выражения для количества заявок в системах массового обслуживания (СМО) вида $M|M|\infty$ в стационарном режиме [11]. Рассмотренные Эрлангом модели по-прежнему являются важными в современной теории массового обслуживания. Однако для приложений важно не только знать стационарное распределение, но и уметь оценивать скорость сходимости к этому распределению. В общем виде эта проблема не решена до сих пор, однако предпринимались некоторые попытки её решения. Так, в случае экспоненциально распределённых времён обслуживания и входящего пуассоновского потока такая проблема рассматривалась в [8, 25] и др.

Известная теорема Б.А. Севастьянова [4, 5] для марковских процессов позволила доказывать не только существование и единственность стационарного распределения для большого класса СМО, но и сходимость в метрике полной вариации. Несколько ранее П. Форте показал [13], что стационарное распределение существует в рассмотренных Б.А. Севастьяновым моделях при несколько более сильных предположениях, чем в работе [5], и указал вид стационарного распределения (без исследования скорости сходимости).

Таким образом, для системы с входящим пуассоновским потоком и произвольным (одинаковым) временем обслуживания с конечным средним значением на всех приборах было известно стационарное распределение; более того, в работах Л. Такача [24] были получены характеристики периода занятости и длины очереди для СМО $M|GI|\infty$.

Однако только в некоторых специальных случаях оценивалась скорость сходимости к стационарному распределению стохастических систем (не обязательно систем Эрланга). Обычно на практике оценки характеристик работы СМО вычисляются для стационарного состояния, но, как правило, в явном виде они не получены, за исключением ряда специальных случаев (см., например, [20]).

Итак, для СМО $M|GI|\infty$ известны стационарные характеристики, но до момента достаточно близкого сближения нестационарных и стационарных значений без знания оценки скорости сходимости распределения состояния СМО к стационарному состоянию невозможно использовать стационарные характеристики. То есть очень важно знать оценку сверху скорости сходимости распределения СМО к стационарному распределению. Более того, если не удаётся определить или оценить стационарное распределение СМО, знание оценки скорости сходимости к нему позволяет получать адекватные оценки характеристик СМО с помощью математического моделирования, а также численно определять стационарные характеристики СМО.

В работах [26, 27, 28, 30] были рассмотрены методы, с помощью которых можно получить оценки скорости сходимости для СМО в случае, когда все распределения абсолютно непрерывны и их интенсивности отделены от нуля и ограничены сверху. Во всех рассмотренных в этих работах моделях предполагалось, что обслуживание начинается сразу в момент поступления требования в систему, что является естественной идеализацией. Также в некоторых случаях предполагалось, что интенсивности (скорости) поступления требований и обслуживания могут зависеть от полного состояния системы.

Задача представленной статьи – построить строгую оценку сверху для скорости сходимости распределения СМО с непуассоновским входящим потоком, возможным запаздыванием начала обслуживания при поступлении требования и отказом от абсолютной непрерывности распределения (зависимых между собой) времён обслуживания.

2. Обобщённая СМО Эрланга – Севастьянова

Итак, мы рассматриваем обобщённую СМО $M|G|\infty$, где интенсивность входящего потока и интенсивности обслуживания зависят от полного состояния системы X_t, которое будет описано ниже. Вообще говоря, при произвольном виде зависимостей изучение поведения такого сорта СМО практически невозможно.

2.1. ПОЛНОЕ СОСТОЯНИЕ СИСТЕМЫ X_T

Полное состояние системы в момент времени t включает:

– $x_t^{(0)}$ – время, прошедшее споследнего поступления требования в СМО;

– $x_t^{(i)}$ – времена нахождения требований, занумерованных по порядку поступления, в системе;

– для удобства добавим n_t – количество требований в СМО в момент t.

То есть полное состояние системы записывается вектором

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

при этом $x_t^{(i)} \geqslant x_t^{(i+1)}$ для $i = 1, \ldots, n_t$, поскольку требования занумерованы в порядке поступления.

Очевидно, $x_t^{(0)} \leqslant x_t^{(n_t)}$, поскольку требование с номером n_t обслуживалось не дольше чем $t - x_t^{(0)}$.

Пространство состояний процесса X_t – это объединение декартовых произведений $\mathcal{X} \stackrel{\text{def}}{=} \bigcup\limits_{n=0}^{\infty} \left(\{n\} \times \mathbb{R}_{\geqslant 0}^{n+1}\right) = \{\vec{\chi}\}$.

Итак, мы предполагаем, что интенсивность входящего потока – это $\lambda = \lambda(X_t)$, а интенсивность обслуживания i-го требования – $h_i = h_i(X_t)$ (рис. 1).

Для упрощения изложения здесь мы полагаем $X_0 = (0; 0)$.

Заметим, что нулевое значение первой n_t компоненты вектора X_t указывает на то, что описываемая СМО свободна – в ней нет требований, т.е. множество свободных состояний процесса обслуживания X_t – это множество

$$\mathcal{S}_0 = \{0\} \times \mathbb{R}_{\geqslant 0} = \{X_t : \, n_t = 0\}.$$

Если же $n_t > 0$, то в системе находится n_t требований. С течением времени периоды, где система свободна, чередуются с периодами занятости, когда в системе находится некоторое (переменное) количество требований. Обозначим свободные периоды σ_i, а периоды занятости – ζ_i.

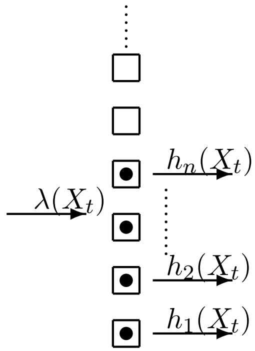

Рис. 1. Входящий поток $\lambda(X_t)$, заняты приборы $1, 2, \dots n$, интенсивность обслуживания на занятых приборах $h_i(X_t)$, $i = 1, 2, \dots n$.

2.1.1. Интенсивность

Напомним, что такое интенсивность.

Пусть длительность некоторого периода ξ, начавшегося в момент $t = 0$, имеет функцию распределения (ф.р.) $F(s) \overset{\mathrm{def}}{=} \mathbb{P}\{\xi \leqslant s\}$, и почти всюду (п.в.) существует $f(s) \overset{\mathrm{def}}{=} F'(s)$. Тогда если к моменту времени s этот период не закончился, то при малых $\Delta > 0$ вероятность того, что он закончится в промежутке $(s, s + \Delta)$, равна

$$\mathbb{P}\{\xi \in (s, s+\Delta) \,|\, \xi > s\} = \frac{\mathbb{P}\{\xi \in (s, s+\Delta) \,\&\, \xi > s\}}{\mathbb{P}\{\xi > s\}} = \\ = \frac{F(s+\Delta) - F(s)}{1 - F(s)} = \frac{f(s + \theta\Delta)\Delta}{1 - F(s)},$$

где $\theta \in [0, 1]$. Поэтому существует предел

$$\lim_{\Delta \downarrow 0} \frac{\mathbb{P}\{\xi \in (s, s+\Delta) \,|\, \xi > s\}}{\Delta} = \frac{f(s)}{1 - F(s)} = I_F(s),$$

который называется интенсивность завершения периода ξ в момент s (естественно, при условии, что до момента s период ξ не завершился).

Несложно заметить, что $F(s)$ удовлетворяет дифференциальному уравнению $\dfrac{F'(s)}{1-F(s)} = I_F(s)$, и

$$F(s) = 1 - \exp\left(-\int_0^s I_F(u)\,\mathrm{d}u\right). \tag{1}$$

Для случая, когда ф.р. $F(s)$ имеет точки разрыва, т.е. распределение случайной величины (сл.в.) ξ имеет дискретную составляющую, то в [35] было замечено, что производная разрывной ф.р. представляется как

$$f(s) = \begin{cases} F'(s), & \text{если существует } F'(s); \\ \\ \delta(0)\big(F(s+0) - F(s-0)\big) & \text{в противном случае,} \end{cases}$$

где $\delta(s)$ – хорошо известная (обобщённая) δ-функция, т.е. $\delta(s) \equiv 0$ при $s \neq 0$ и $\int\limits_{-a}^{a} \delta(s)\,\mathrm{d}s = 1$ для всех $a > 0$.

Так что можно определить *обобщённую интенсивность* периода ξ следующим образом:

$$I_F(s) \stackrel{\text{def}}{=} \frac{f(s)}{1-F(s)} - \sum_i \delta(s-a_i)\ln\big(F(a_i+0) - F(a_i-0)\big),$$

где $\{a_i\}$ — множество точек разрыва ф.р. $F(s)$.

При этом несложно убедиться, что соотношение (1) остаётся в силе.

Таким образом, неотрицательные (смешанные) случайные величины можно задавать как с помощью ф.р. и плотности, так и с помощью интенсивности.

***<u>Замечание</u>* 1.** Для того чтобы неотрицательная (обобщённая) функция $\mu(s)$, заданная для $s \geqslant 0$, задавала распределение собственной сл.в., необходимо и достаточно, чтобы

$$\int_0^\infty \mu(s)\,\mathrm{d}s = +\infty.$$

Напомним, что сл.в. является собственной, если она с вероятностью 1 меньше бесконечности.

Итак, входящий поток имеет (переменную) обобщённую интенсивность $\lambda(X_t)$, времена обслуживания имеют (переменные) обобщённые интенсивности $h_i(X_t)$, и все эти интенсивности зависят от полного состояния СМО X_t; при этом *весь процесс* X_t является марковским процессом на пространстве состояний $\mathcal{X}$; его переходные вероятности определяются функциями $\lambda(X_t)$ и $h_i(X_t)$.

Очевидно, что абсолютно произвольная зависимость $\lambda(X_t)$ и $h_i(X_t)$ от состояния процесса X_t не позволяет сделать какие-либо выводы о поведении и распределении процесса X_t.

Поэтому далее вводятся условия, при которых рассматривается поведение процесса X_t.

2.2. УСЛОВИЯ НА (ОБОБЩЁННЫЕ) ИНТЕНСИВНОСТИ

1. Существует п.в. неотрицательная функция $\lambda_0(s)$ и постоянная Λ такие, что для всех возможных значений X_t выполнено условие: $\lambda_0\left(x_t^{(0)}\right) \leqslant \lambda(X_t) \leqslant \Lambda < \infty$.

2. $\mathfrak{M}_k \stackrel{\text{def}}{=} \int\limits_0^\infty s^k \, \mathrm{d}\mathfrak{L}(s) < \infty$ для $k \geqslant 2$, где

$$\mathfrak{L}(s) = 1 - \exp\left(-\int\limits_0^s \lambda_0(s)\,\mathrm{d}s\right) \mathfrak{L}(+\infty) = 1.$$

3. Существует постоянная $T \geqslant 0$ и существуют две неотрицательные (обобщённые) функции $\varphi(t)$ и $q(t)$, причём при всех $t > T$ п.в. $\varphi(t) > 0$; кроме того, $q(t) \neq \delta(0)$.

Интенсивности $h_i(X_t)$ $(i = 1, 2, \ldots n_t)$ удовлетворяют условию $\varphi\left(x_t^{(i)}\right) \leqslant h_i(X_t) \leqslant q\left(x_t^{(i)}\right)$ для всех возможных значений $X_t = \left(n_t; x_t^{(0)}; x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n_t)}\right)$.

4. $\int\limits_0^\infty s^k \, \mathrm{d}\Phi(s) < \infty$ для $k \geqslant 2$, где

$$\Phi(s) = 1 - \exp\left(-\int\limits_0^s \varphi(u)\, \mathrm{d}u\right) \Phi(+\infty) = 1.$$

Замечание 2. Условия 1 и 2 гарантируют, что входящий поток требований имеет, возможно, переменную интенсивность, но в любом случае эта интенсивность является необобщённой функцией и не превосходит величины Λ, и математическое ожидание времени между поступлениями требований имеет не менее k конечных моментов.

Замечание 3. Условие 3 показывает:

– Время запаздывания обслуживания (т.е. время от момента поступления в СМО до начала обслуживания) не превосходит величины T (на промежутке $(0, T)$ интенсивность обслуживания может быть нулевой).

– Функция распределения (ф.р.) времени пребывания в системе (запаздывание + обслуживание) любого требования заключена в границах $\Phi(s) \leqslant F_j(s) \leqslant Q(s)$, где

$$Q(s) \stackrel{\mathrm{def}}{=} 1 - \exp\left(-\int\limits_0^s q(u)\, \mathrm{d}u\right),$$

а $F_j(s)$ – ф.р. времени обслуживания j-го по порядку поступления в систему требования ς_j.

– При этом $\int\limits_0^\infty s^k \, \mathrm{d}F_j(s) \leqslant \int\limits_0^\infty s^k \, \mathrm{d}\Phi(s) < \infty$, т.е. существует не менее k конечных моментов у всех времён обслуживания.

– Ф.р. $F_i(s) = 1 - \exp\left(-\int\limits_0^s \lambda(u)\, \mathrm{d}u\right)$ времени между поступлениями требований $\hat{\sigma}_i$ заключена в границах

$$\mathfrak{L}(s) \leqslant F_i(s) \leqslant \widetilde{\mathfrak{L}}(s), \text{ где } \widetilde{\mathfrak{L}}(s) \stackrel{\text{def}}{=} 1 - \exp\left(-\int\limits_0^s \Lambda(s)\,\mathrm{d}s\right).$$

Соответственно, плотность распределения $F_i'(s)$ времени между поступлениями требований $\hat{\sigma}_i$ оценивается как $\lambda_0 \exp\left(-\int\limits_0^s \Lambda(s)\,\mathrm{d}s\right) \leqslant F_i'(s) \leqslant \Lambda \exp\left(-\int\limits_0^s \lambda_0(s)\,\mathrm{d}s\right)$.

Из условия $q(t) \neq \delta(0)$ следует, что все (существующие) моменты всех времён обслуживания положительны.

***Замечание* 4.** Условие 4 гарантирует наличие не менее k конечных моментов у времени пребывания требований в системе.

Наша цель – доказать эргодичность процесса X_t и оценить сверху скорость сходимости его распределения к предельному.

Для этого процесс сначала сравним с хорошо исследованным процессом обслуживания «классической» СМО Эрланга – Севастьянова $M|G|\infty$ и получим предварительные оценки характеристик исследуемой СМО. Будет доказана эргодичность процесса обслуживания исследуемой СМО. Затем с использованием метода склеивания будет представлен алгоритм получения строгих оценок скорости сходимости распределения этой СМО к стационарному распределению.

3. Эргодичность процесса X_t

Напомним, что период занятости СМО – это период, который начинается, когда требование приходит в систему, в которой нет других требований, и заканчивается, когда (это же или другое) требование покидает систему, в которой больше нет требований; причём на всём этом периоде в СМО присутствует по крайней мере одно требование.

Отметим, что X_t – это регенерирующий процесс: моменты регенерации – это моменты времени, когда X_t попадает в состояние $X_t = (1;0;0)$, т.е. до этого времени система была свободна ($n_t = 0$), и в этот момент в свободную систему приходит требование.

Обозначим R_1, R_2, R_3, $\ldots$ – длины последовательных периодов регенерации. Эти сл.в. являются независимыми и одинаково распределёнными (н.о.р.), они состоят из свободного периода (или периода простоя) σ_i и периода занятости ζ_i.

Для эргодичности процесса X_t достаточно ограниченности математического ожидания периодов регенерации:

$$\sup_{i\in\mathbb{N}} \mathbb{E}(\sigma_i + \zeta_i) < C < \infty \ \Rightarrow X_t \Longrightarrow \widetilde{X}.$$

В этом случае $\lim\limits_{t\to\infty} \mathcal{P}_t(A) = \widetilde{\mathcal{P}}_t$ для любых измеримых A из пространства состояний $\mathcal{F}$ процесса X_t:

$$\mathcal{F} \stackrel{\mathrm{def}}{=} \mathbb{R}_+ \cup (\mathbb{R}_+ \times \mathbb{R}_+) \cup \mathbb{R}_+^3 \cup \dots \cup \mathbb{R}_+^n \cup \dots = \bigcup_{i=1}^{\infty} \mathbb{R}_+^i,$$

а $\mathcal{P}_t$ – распределение процесса X_t: $\mathcal{P}_t(A) = \mathbb{P}\{X_t \in A\}$; $\widetilde{\mathcal{P}}$ – стационарное распределение этого процесса, инвариантное относительно переходных вероятностей, задаваемых интенсивностями $\lambda(\widetilde{X})$, $h_i(\widetilde{X})$.

3.1. ПЕРИОД ПРОСТОЯ ПРОЦЕССА X_T

Период простоя σ_i не превосходит интервала $\widehat{\sigma}_i$ – времени между поступлением $(i-1)$-го и i-го требований в систему ($\widehat{\sigma}_1 = \sigma_1$ – момент поступления 1-го требования в систему).

Из условия 2 раздела 2.2 имеем для всех $\ell \in [0, k]$

$$\mathbb{E}\,\sigma_i^\ell \leqslant \mathbb{E}\,\widehat{\sigma}_i^\ell \leqslant \int\limits_0^\infty s^\ell \,\mathrm{d}\mathfrak{L}(s) = \mathfrak{M}_\ell. \tag{2}$$

3.2. ПЕРИОД ЗАНЯТОСТИ ПРОЦЕССА X_T

Оценим моменты сл.в. ζ_i. Для этого сравним поведение процесса X_t с уже изученным во многих работах процессом обслуживания СМО $M|G|\infty$ (см., например, [24]).

3.3. СРАВНЕНИЕ ПРОЦЕССА X_T СО «СТАНДАРТНЫМ» ПРОЦЕССОМ ДЛЯ $M|G|\infty$

Рассмотрим «стандартную» систему массового обслуживания $M|G|\infty$, в которой входящий поток имеет постоянную интенсивность Λ, т.е. времена между поступлениями требований независимы и имеют ф.р. $\mathcal{L}(s) \stackrel{\text{def}}{=} 1 - e^{-\Lambda s}$.

Обозначим σ_i^Y – свободные периоды процесса Y_t, а ζ_i^Y – его периоды занятости.

Времена обслуживания ζ_j^Y поступающих в систему независимы и имеют одинаковую ф.р.

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

(Λ и $\varphi(s)$ – из условий раздела 2.2).

Времена между поступлением заявок и времена обслуживания независимы в совокупности.

Как и в рассматриваемой обобщённой СМО Эрланга – Севастьянова, мы строим марковский процесс $Y_t = \left(m_t; y_t^{(0)}; y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(n_t)}\right)$, где m_t – количество требований в системе в момент t, $y_t^{(0)}$ – время, прошедшее с момента последнего поступления требования, а $y_t^{(i)}$ – это время, в течение которого i-е требование находилось в системе до момента t; $Y_0 = (0; 0)$.

Определение 1. *Мы говорим, что $\xi \prec \eta$, если $\mathbb{P}\{\xi < s\} \geqslant \mathbb{P}\{\eta < s\}$ для всех $s \in \mathbb{R}$.*

Отношение $\prec$ – это отношение порядка на множестве случайных величин.

Заметим, что для любых номеров i и j выполнены соотношения $\sigma_i^Y \prec \sigma_i$ и $\zeta_i \prec \zeta_i^Y$, т.е. в процессе Y_t требования приходят чаще, а обслуживаются дольше, чем в процессе X_t.

Лемма 1. *Можно построить процессы X_t и Y_t на одном и том же вероятностном пространстве таким образом, что:*

1) $X_0 = Y_0 = (0; 0)$;

2) $n_t \leqslant m_t$ *для всех* $t \geqslant 0$.

Доказательство основано на фактах из [23]; см. также [2].

Поэтому период занятости ζ^X для X_t не превосходит периода занятости ζ^Y для Y_t в смысле определения 1: $\zeta^X \prec \zeta^Y$; более того, $\mathbb{P}\{n_t = 0\} \geqslant \mathbb{P}\{m_t = 0\}$, и $\mathbb{E}\left[\zeta^X\right]^k \leqslant \mathbb{E}\left[\zeta^Y\right]^k$ для всех $k > 0$.

3.3.1. Хорошо известные факты о «стандартной» системе $M|G|\infty$

В «стандартной» системе $M|G|\infty$, поведение которой описывается процессом Y_t, интенсивность входящего потока – постоянная Λ, времена обслуживания ξ_i^Y являются н.о.р.сл.в. с ф.р. $\mathbb{P}\{\xi_i^Y < s\} = \Phi(s) = 1 - \exp\left(-\int\limits_0^s \varphi(t)\,\mathrm{d}t\right)$.

Для этой системы (при $X_0 = (0;0)$) [12, 22, 24]:

– $\mathbb{P}\{m_t = k\} = \dfrac{e^{-\mathcal{G}(t)}\big(\mathcal{G}(t)\big)^k}{k!}$, где $\mathcal{G}(t) \stackrel{\text{def}}{=} \Lambda \int\limits_0^t (1 - \Phi(s))\,\mathrm{d}s$.

То есть для всех $t \geqslant 0$ верно

$$\mathbb{P}\{m_t = 0\} \geqslant \lim_{t\to\infty} \mathbb{P}\{m_t = 0\} = e^{-\rho}, \tag{3}$$

где

$$\rho \stackrel{\text{def}}{=} \Lambda m_1 = \Lambda \int\limits_0^\infty (1 - \Phi(s))\,\mathrm{d}s. \tag{4}$$

– Для периода занятости рассматриваемой системы $M|G|\infty$ (это периоды занятости ζ_i^Y) известна функция распределения

$$B(s) \stackrel{\text{def}}{=} \mathbb{P}\{\zeta_i^Y \leqslant s\} = 1 - \frac{1}{\Lambda}\sum_{k=1}^{\infty} c^{n*}(s), \tag{5}$$

где $c(s) \stackrel{\text{def}}{=} \Lambda(1-\Phi(s))e^{-\mathcal{G}(s)}$, и $c^{n*}(s)$ явлется n-й свёрткой функции $c(s)$.

– Кроме того, мы знаем преобразование Лапласа для $B(s)$:

$$\mathcal{L}[B](s) = 1 + \frac{s}{\Lambda} - \frac{1}{\Lambda \int\limits_0^\infty \exp\left(-st - \Lambda \int\limits_0^t [1 - \Phi(v)]\,\mathrm{d}v\right)\,\mathrm{d}t}. \tag{6}$$

– Также известны формулы:

(7) $$\mathbb{E}\,\left(\zeta_i^Y\right)^n = \\ = (-1)^{n+1}\left\{\frac{e^{\rho}\,n\,C^{(n-1)}}{\Lambda} - e^{\rho}\sum_{k=1}^{n-1} C_n^k\,\mathbb{E}\,(\zeta_i^Y)^{n-k}C^{(p)}\right\}, \\ n \in \mathbf{N},$$

где

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

Из формул (5)–(7) несложно получить

$$\mathbb{E}\,\zeta_i^Y = \frac{e^{\rho}-1}{\Lambda} < \infty.$$

Следовательно, период регенерации процесса X_t, который представляет собой сумму $\sigma_i+\zeta_i$, имеет конечное математическое ожидание:

$$\mathbb{E}(\sigma_i+\zeta_i) \leqslant \mathfrak{M}_1 + \frac{e^{\rho}-1}{\Lambda} < \infty.$$

Таким образом, доказана

Теорема 1. *В условиях 1–4 (раздел 2.2) процесс X_t эргодичен, т.е. распределение $\mathcal{P}_t$ процесса X_t слабо сходится к стационарному инвариантному распределению $\widetilde{\mathcal{P}}$.*

4. Скорость сходимости распределения процесса X к стационарному распределению

Для того чтобы оценить скорость сходимости $\mathcal{P}_t \Longrightarrow \widetilde{\mathcal{P}}$, достаточно уметь оценивать моменты периода регенерации процесса X_t.

Как было показано в (2), для свободных периодов σ_i верно неравенство $\mathbb{E}\,\sigma_i^{\ell} \leqslant \mathfrak{M}_{\ell}$ для $\ell \in [0,k]$.

Из формул (6) и (7) выше была получена формула

$$\mathbb{E}\,\zeta_i \leqslant \mathbb{E}\,\zeta_i^Y = \frac{e^{\rho}-1}{\Lambda},$$

также несложно показать, что

$$\mathbb{E}\,(\zeta_i)^2 \leqslant \mathbb{E}\,(\zeta_i^Y)^2 = \frac{2e^{2\rho}}{\Lambda}\int\limits_0^\infty \exp\left(-\Lambda\int\limits_0^\infty [1-\Phi(v)]\,\mathrm{d}v\right)\,\mathrm{d}t.$$

Но для $\ell > 2$ вычисления моментов $\mathbb{E}\,(\zeta_i^Y)^\ell$ очень сложны, поэтому мы хотим оценить эти моменты.

4.1. ОЦЕНКА МОМЕНТОВ ζ_I^Y

Вернёмся к формуле $B(s) \stackrel{\text{def}}{=} \mathbb{P}\{\zeta_i^Y \leqslant s\} = 1 - \dfrac{1}{\Lambda}\sum\limits_{k=1}^{\infty} c^{n*}(s)$, где $c(s) \stackrel{\text{def}}{=} \Lambda(1-\Phi(s))e^{-\mathcal{G}(s)}$, и c^{n*} – n-я свёртка функции $c(s)$.

Легко видеть, что

$$\int\limits_0^\infty c(s)\,\mathrm{d}s = 1 - e^{-\Lambda m_1} = 1 - e^{-\rho} \stackrel{\text{def}}{=} \varrho \in (0;1).$$

Поэтому $r(s) \stackrel{\text{def}}{=} \dfrac{c(s)}{\varrho}$ – это плотность распределения (п.р.) некоторой сл.в. ς.

Заметим, что $c^{n*}(s) = \varrho^n r_n(s)$, где $r_n(s)$ – п.р. суммы $\sum\limits_{i=1}^{n}\varsigma_i$, и ς_i – это н.о.р.сл.в. с п.р. $r(s)$.

Далее,

$$\mathbb{E}\,\varsigma^k = \int\limits_0^\infty s^k\varrho^{-1}\Lambda(1-\Phi(s))e^{-\mathcal{G}(s)}\,\mathrm{d}s \leqslant$$

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

Теперь, применяя неравенство Йенсена в виде

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

для неотрицательных $a_1, \ldots, a_n$ и $\ell \leqslant k$ (см. условия раздела 2.2), получаем:

$$\mathbb{E}\,(\zeta_i^Y)^\ell = \int\limits_0^\infty \ell s^{\ell-1}(1-B(s))\,\mathrm{d}s = \frac{1}{\Lambda}\sum_{n=1}^{\infty}\int\limits_0^\infty \ell s^{\ell-1}c^{n*}(s)\,\mathrm{d}s =$$

$$= \frac{1}{\Lambda}\sum_{n=1}^{\infty}\int\limits_0^\infty \ell s^{\ell-1}\varrho^n r^{n*}(s)\,\mathrm{d}s = \frac{1}{\Lambda}\sum_{n=1}^{\infty}\varrho^n \ell\,\mathbb{E}\left(\sum_{j=1}^{n}\varsigma_j\right)^{\ell-1} \leqslant$$

$$\leqslant \frac{1}{\Lambda}\sum_{n=1}^{\infty}\varrho^n \ell n^{\ell-1}\frac{\Lambda\mathbb{E}\,(\xi_i^Y)^\ell}{\varrho\ell} = \frac{\mathbb{E}\,(\xi_i^Y)^\ell}{\varrho}\sum_{n=1}^{\infty}n^{\ell-1}\varrho^n =$$

$$= \frac{\mathbb{E}\,(\xi_i^Y)^\ell}{\varrho}\times\varphi(\varrho,\ell-1),$$

где $\varphi(s,\ell) \stackrel{\mathrm{def}}{=\!=} \left(s\,\dfrac{\mathrm{d}}{\mathrm{d}s}\right)^\ell \dfrac{1}{1-s}$, и, в частности,

$$\mathbb{E}\,\zeta_i^Y \leqslant \frac{e^\rho - 1}{\Lambda}. \tag{8}$$

4.1.1. Период регенерации процесса X_t

Поскольку период регенерации – это сумма двух (зависимых) сл.в. $\sigma_i + \zeta_i \leqslant \sigma_i^Y + \zeta_i^Y$, и их м.о. ограничены (см. (2), (8)), то длина периода регенерации ограничена.

Кроме того, поскольку выполнены условия 1–4 из раздела 2.2 мы выводим, что $\mathbb{E}\,(\sigma_i + \zeta_i) \leqslant \mathbb{E}\,(\sigma_i^Y + \zeta_i^Y) < \infty$. Т.е. процесс X_t регенерирующий и эргодический. Следовательно, его распределение $\mathcal{P}_t$ в момент t сходится к инвариантному предельному распределению $\mathcal{P}$: $\mathcal{P}_t \Longrightarrow \mathcal{P}$. Итак, процесс X_t эргодичен.

Построим оценку сверху для скорости сходимости распределения $\mathcal{P}$ процесса X_t к предельному распределению. Для вычисления оценки скорости сходимости $\mathcal{P}$: $\mathcal{P}_t \Longrightarrow \mathcal{P}$ будет использоваться метод склеивания. Этот метод хорошо известен в зарубежной литературе, но в русскоязычных изданиях информация о нём крайне отрывочна (см., например, [1]).

При этом многие отечественные исследователи использовали и развивали этот метод склеивания в англоязычных публикациях (см., например, [15]).

4.2. МЕТОД СКЛЕИВАНИЯ

Принцип использования метода склеивания заключается в следующем.

Если два однородных марковских процесса X_t и X'_t с одной и той же переходной функцией (переходными вероятностями), но с различными начальными состояниями совпали в момент τ, то после момента τ их распределения одинаковы.

Поэтому выполнено *основное неравенство склеивания*

$$(9)\quad |\mathbb{P}\{X_t \in S\} - \mathbb{P}\{X'_t \in S\}| = \\ = |\mathbb{P}\{X_t \in S\} - \mathbb{P}\{X'_t \in S\}| \times (\mathbf{1}(\tau > t) + \mathbf{1}(\tau \leqslant t)) \leqslant \mathbb{P}\{\tau > t\},$$

и для любой положительной неубывающей функции $\phi(s)$ верно

$$|\mathbb{P}\{X_t \in S\} - \mathbb{P}\{X'_t \in S\}| \leqslant \mathbb{P}\{\tau > t\} = \\ = \mathbb{P}\{\phi(\tau) > \phi(t)\} \leqslant \frac{\mathbb{E}\,\phi(\tau)}{\phi(t)}.$$

(В данной работе мы будем рассматривать функцию $\phi(s) = s^\ell$).

Для цепей Маркова (т.е. процессов Маркова в дискретном времени) метод склеивания впервые был предложен в [10].

Для марковских процессов в непрерывном времени в [14] была предложена следующая модификация метода склеивания.

4.2.1. Параллельное склеивание (“successful coupling”)

Для реализации этого приёма для пары двух однородных марковских процессов X_t и X'_t с одинаковыми переходными функциями, но разными начальными состояниями, *на некотором вероятностном пространстве* конструируется «параллельная» пара процессов $\mathcal{Z}_t = (Z_t, Z'_t)$ такая, что

(i) Для всех $s \geqslant 0$ и $S \in \mathcal{B}(\mathcal{X})$ выполнены равенства $\mathbf{P}\{Z_t \in S\} = \mathbf{P}\{X_t \in S\}$, $\mathbf{P}\{Z'_t \in S\} = \mathbf{P}\{X'_t \in S\}$. (Соответственно, $Z_0 = X_0$ и $Z'_0 = X'_0$.)

(ii) Для всех

$$t > \tau(X_0, X'_0) = \tau(Z_0, Z'_0) \stackrel{\text{def}}{=} \inf\{t \geqslant 0 : \, Z_t = Z'_t\}$$

выполняется равенство $Z_t = Z'_t$.

(iii) Для всех возможных начальных состояний X_0, X'_0 процессов X_t и X'_t из пространства состояний $\mathcal{X}$ верно равенство $\mathbf{P}\{\tau(X_0, X'_0) < \infty\} = 1$.

Если условия (i)–(iii) выполнены, то для парного («параллельного» паре (X_t, X'_t)) процесса $\mathcal{Z}_t$ момент $\tau = \tau(X_0, X'_0)$ назовём успешной параллельной склейкой [14]. *Основное неравенство склеивания* (9) преобразуется так:

$$|\mathbb{P}\{X_t \in S\} - \mathbb{P}\{X'_t \in S\}| = |\mathbb{P}\{Z_t \in S\} - \mathbb{P}\{Z'_t \in S\}| \leqslant$$

$$\leqslant \mathbb{P}\{\tau > t\} \leqslant \mathbb{P}\{\phi(\tau) > \phi(t)\} \leqslant \frac{\mathbb{E}\,\phi(\tau)}{\phi(\tau)} = \mathcal{R}(t, X_0, X'_0)$$

для неубывающей положительной функции $\phi(s)$.

Поэтому

$$\|\mathcal{P}_t - \mathcal{P}'_t\|_{TV} \stackrel{\text{def}}{=} 2 \sup_{S \in \mathcal{B}(\mathcal{X})} |\mathbb{P}\{X_t \in S\} - \mathbb{P}\{X'_t \in S\}| \leqslant$$

$$\leqslant 2\mathcal{R}(t, X_0, X'_0).$$

4.2.2. Основная лемма склеивания

Эта лемма (Basic Coupling Lemma – см., например, [17, 29]) применяется для непрерывных распределений.

***Лемма* 2** Основная лемма склеивания. *Пусть $f_i(s)$ – плотность распределения сл.в. θ_i ($i = 1,2$). И предположим, что $\int\limits_{-\infty}^{\infty} \min(f_1(s), f_2(s))\,\mathrm{d}s = \widetilde{\varkappa} > 0$. Тогда на некотором вероятностном пространствесуществуют две сл.в. ϑ_i такие, что $\vartheta_i \stackrel{\mathcal{D}}{=} \theta_i$, и $\mathbb{P}\{\vartheta_1 = \vartheta_2\} \geqslant \widetilde{\varkappa}$.*

Доказательство. Оно является чисто конструктивным – см., например, [17] и [9]. Приведём простейший вариант.

Пусть $\mathcal{U}_i$ независимые равномерно распределённые на $(0;1)$ сл.в.. Положим

$$\phi(s) \stackrel{\text{def}}{=} \widetilde{\varkappa}^{-1} \int\limits_{-\infty}^{s} \min\big(f_1(s), f_2(s)\big)\,\mathrm{d}s,$$

$$\psi_i \stackrel{\text{def}}{=} (1-\widetilde{\varkappa})^{-1} \left(\int\limits_{-\infty}^{s} f_i(s) - \min\big(f_1(s), f_2(s)\big)\,\mathrm{d}s \right).$$

Положим $\vartheta_i \stackrel{\text{def}}{=} \mathbf{1}(\mathcal{U}_0 < \widetilde{\varkappa})\phi^{-1}(\mathcal{U}_4) + \mathbf{1}(\mathcal{U}_0 \geqslant \widetilde{\varkappa})\psi_i^{-1}(\mathcal{U}_i)$.

Легко видеть, что $\mathbb{P}\{\vartheta_1 = \vartheta_2\} = \widetilde{\varkappa}$.

***Замечание* 5.** Величина $\widetilde{\varkappa}$ называется общей частью распределений $f_1(s)$ и $f_2(s)$.

5. Применение метода «параллельного склеивания» к процессу X_t

Вернёмся к изучаемому процессу X_t (см. Раздел 2) и его версии X'_t (с другим начальным состоянием).

Для простоты будем считать, что $X_0 = (0;0)$, а начальное состояние X'_0 – произвольное из пространства состояний $\mathcal{X}$;

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

$$X'_t = \left(n'_t, x_t^{(0)'}; x_t^{(1)'}, x_t^{(2)'}, \ldots, x_t^{(n'_t)'}\right).$$

Наша задача – построить такие «параллельные» копии процессов X_t и X'_t, чтобы можно было оценить момент склеивания этих копий; понятно, что этот момент склеивания зависит от X'_0. Обозначим его $\tau(X'_0)$.

Конструирование пары «параллельных» процессов (Z_t, Z'_t), имеющих такие же маргинальные распределения, что и пара (X_t, X'_t), будем строить, изменяя поведение процессов X_t и X'_t в выбранные марковские моменты времени таким образом, что продолжение их «траекторий» с некоторой вероятностью совпадёт в следующий марковский момент.

Рассмотрим моменты времени θ_1, θ_2, θ_3,... когда X'_t попадает в множество $\mathcal{S}_0$, т.е. n'_t переходит из значения 1 в значение 0.

Из формул (3)–(4) имеем:

$$\mathbb{P}\{X_{\theta_i} \in \mathcal{S}_0\} = \mathbb{P}\{n_{\theta_i} = 0\} \geqslant \mathbb{P}\{m_{\theta_i} = 0\} \geqslant e^{-\rho} \stackrel{\text{def}}{=} \pi_0. \tag{10}$$

(Напомним, что m_t – число требований во вспомогательном процессе Y_t.) Если $n_{\theta_i+} = n'_{\theta_i+}$, то оба процесса X_t и X'_t оказываются в свободном состоянии; вероятность такого события не менее числа π_0.

Заметим, что свободные периоды процесса X_t являются *процессами почти-восстановления*, и к ним применимо *обобщённое неравенство Лордена* (см. [16, 35]).

5.1. ПРОЦЕССЫ ПОЧТИ-ВОССТАНОВЛЕНИЯ И ОБОБЩЁННОЕ НЕРАВЕНСТВО ЛОРДЕНА

Рассмотрим считающий процесс со скачками $N_t \stackrel{\text{def}}{=} \sum\limits_{i=1}^{\infty} \mathbf{1}\left\{\sum\limits_{k=1}^{i} \xi_k \leqslant t\right\}$, где $\{\xi_1, \xi_2, ...\}$ – независимые одинаково распределённые (н.о.р.) положительные случайные величины. Рассмотрим перескок (или обратное время восстановления) для процесса N_t: $b_t = t - \sum\limits_{k=1}^{N_t} \xi_k$ – это время от момента t до следующего скачка процесса N_t.

Теперь предположим, что в этом считающем процессе N_t сл.в. ξ_j могут не быть н.о.р.. Очевидно, при совершенно произвольной зависимости ξ_j и ξ_i и совершенно произвольных распределениях ξ_j или ξ_i ничего существенного о поведении такого процесса N_t сказать нельзя.

Предполагаем, что $\mathbb{P}\{\xi_j \leqslant s\} = F_j(s)$; F_j и F_i могут быть различными, но выполнены условия 1–3; $\Phi(x) \geqslant F_j(x) \geqslant Q(x)$ – см. выше раздел 2.2.

Определение 2. Считающий процесс

$$N_t \stackrel{\text{def}}{=} \sum_{i=1}^{\infty} \mathbf{1}\left\{\sum_{k=1}^{i} \xi_k \leqslant t\right\},$$

где $\{\xi_1, \xi_2, ...\}$ удовлетворяют условиям 1–3 раздела 2.2, называется процессом почти-восстановления.

Замечание 6. Ранее другое обобщение процессов восстановления рассматривалось в работах [3, 6] и др., и такие процессы назывались *обобщёнными процессами восстановления*.

Замечание 7. Заметим, что «в чистом виде» процессы почти-восстановления рассматривать не имеет смысла: зависимость и неодинаковая распределённость периодов почти-восстановления возникает в реальных ситуациях из-за воздействия внешних параметров – или в тех случаях, когда несколько процессов восстановления взаимодействуют между собой и оказывают друг на друга влияние. В частности, в рассматриваемой в данной работе модели обобщённой СМО Эрланга – Севастьянова периоды между поступлениями требований представляют собой процессы почти-восстановления.

В работе [16] (см. также [35]) приводится следующее обобщение неравенства Лордена ([18]):

Теорема 2. *Если условия 1–3 для интенсивностей сл.в. ξ_i выполнены, и $\mathbb{E}\,\eta^k < \infty$, то для процесса B_t верны следующие неравенства:*

$$\mathbb{E}\,(b_t)^{k-1} \leqslant \mathbb{E}\,\eta^{k-1} + \frac{\mathbb{E}\,\eta^k}{k\mathbb{E}\,\zeta}, \tag{11}$$

где $\mathbb{E}\,\eta^k = \int\limits_0^{\infty} x^k \,\mathrm{d}\Phi(x)$; $\mathbb{E}\,\zeta = \int\limits_0^{\infty} x \,\mathrm{d}Q(x)$;

и $Q(x) = 1 - \int\limits_0^x \exp\left(-\int\limits_0^s q(t)\,\mathrm{d}t\right)\mathrm{d}s$ – *см. обозначения в замечании 3 раздела 2.2.*

***Следствие* 1.** *При выполнении условий 1–3 раздела 2.2 верно*

$$\mathbb{E}\,(b_t) \leqslant \mathbb{E}\,\eta + \frac{\mathbb{E}\,\eta^2}{2\mathbb{E}\,\zeta}. \tag{12}$$

5.2. ПРОДОЛЖЕНИЕ ПРИМЕНЕНИЯ МЕТОДА «ПАРАЛЛЕЛЬНОГО СКЛЕИВАНИЯ» К ПРОЦЕССУ X_T

Итак, если в некоторый момент времени $\widetilde{\tau}$ один из двух процессов X_t, X'_t попадает в множество $\mathcal{S}_0 \stackrel{\mathrm{def}}{=} \{(0;\alpha)\}$, то с вероятностью большей, чем $\pi_0 = \rho$ (см. (10)), второй процесс находится в свободном состоянии (в соответствующей системе отсутствуют требования), т.е. в этот момент времени. например, $X_{\widetilde{\tau}} = (0;\alpha)$ и $X'_{\widetilde{\tau}} = (0;\beta)$.

В момент $\widetilde{\tau}$ перескоки обоих процессов поступления требований в обе системы могут быть оценены с помощью обобщённого неравенства Лордена. Действительно, в соответствии с (11)–(12)), $\mathbb{E}\,\alpha \leqslant \mathbb{E}\,\eta + \dfrac{\mathbb{E}\,\eta^2}{2\mathbb{E}\,\zeta} \stackrel{\mathrm{def}}{=} \Theta_0$ и $\mathbb{E}\,\beta \leqslant \mathbb{E}\,\eta + \dfrac{\mathbb{E}\,\eta^2}{2\mathbb{E}\,\zeta} \stackrel{\mathrm{def}}{=} \Theta_0$.

Поэтому для всех $\Theta > \Theta_0$ верно $\mathbb{P}\{\alpha < \Theta\} \geqslant 1 - \dfrac{\Theta_0}{\Theta}$ и $\mathbb{P}\{\beta < \Theta\} \geqslant 1 - \dfrac{\Theta_0}{\Theta} \stackrel{\mathrm{def}}{=} \pi_1$, что следует из неравенства Маркова.

Зафиксируем некоторое число $\Theta > \Theta_0$ – от его выбора зависят дальнейшие вычисления.

Обе нижние границы интенсивностей $\varphi(s) > 0$ при $s > T$ для $X_t = \left(0, x_t^{(0)}\right)$, $X'_t = \left(0, {x'}_t^{(0)}\right)$, поэтому общая часть распределений остаточных времён пребывания обоих процессов в свободном (без требований) состоянии не менее, чем:

$$\varkappa(\Theta) \stackrel{\mathrm{def}}{=} \inf_{\alpha<\Theta,\beta<\Theta} \int\limits_0^\infty \min\{\lambda_0(\alpha+s), \lambda_0(\beta+s)\}\,\mathrm{d}s \stackrel{\mathrm{def}}{=} \pi_2.$$

То есть в этом случае можно применить основную лемму склеивания, и в случае, если $X_{\widetilde{\tau}} = (0,\alpha)$ и $X'_{\widetilde{\tau}} = (0,\beta)$, с вероятностью $\mathbb{P}\{\alpha < \Theta \;\&\; \beta < \Theta\} \geqslant \pi_1^2$ можно применять основную лемму склеивания с учётом общей части остаточных времён пребывания в свободном состоянии обеих СМО не менее $\varkappa(\Theta) = \pi_2$, и с вероятностью не меньшей, чем π_2, процессы совпадут в начале следующего периода занятости (т.е. их свободные периоды закончатся одновременно).

Стало быть, с вероятностью большей, чем $\pi \stackrel{\text{def}}{=\!=} \pi_0\pi_1^2\pi_2$, мы можем успешно применить основную лемму склеивания в момент попадания каждого из процессов X_t, X'_t в свободное состояние (естественно, при условии, что рассматриваются полные периоды почти-восстановления). (Полный период почти-восстановления начинается с нулевого значения перескока, т.е. в нашем случае это предыдущее поступление требования в систему.) И, в случае совпадения этих процессов после её применения, их распределения в дальнейшем будут совпадать в любой последующий момент времени.

Иначе говоря, после любого попадания процесса X_t или X'_t в множество $\mathcal{S}_0$ мы с некоторой оцененной вероятностью можем получить момент склеивания процессов X'_t и X_t – в момент перехода из свободного состояния в занятое, т.е. в момент регенерации, когда состояние системы есть $(1, 0; 0)$.

Таким образом, в конце любого периода регенерации процессы X_t и X'_t могут совпасть с вероятностью большей, чем π (эта оценка вероятности может быть улучшена, например, подходящим выбором величины Θ). Следовательно, время склейки $\tau(X'_0)$ – это геометрическая сумма периодов регенерации процесса X'_t включая первый (неполный) период регенерации. Напомним, что $X_0 = (0, 0)$, а X'_0 – произвольный элемент пространства состояний $\mathcal{X}$.

Итак, теперь мы можем найти оценку сверху для $\mathbb{E}\,[\tau(X'_0)]^k$ в случае, когда $\mathbb{E}\,\xi_i^k < \infty$ $(C > k - 1)$:

$$\mathbb{E}\,[\tau(X_0')]^k \leqslant \pi \sum_{i=0}^{\infty}(1-\pi)^i \mathbb{E}\,\left(R_0 + \sum_{j=1}^{i} b_j\right)^k \leqslant$$

$$\leqslant \sum_{i=0}^{\infty}(1-\pi)^i (i+1)^{k-1} \mathbb{E}\,\left(R_0^k + \sum_{j=1}^{i} R_j^k\right) =$$

$$= \mathbb{E}\,R_0^k \times \sum_{i=0}^{\infty}(1-\pi)^i (i+1)^{k-1} + \mathbb{E}\,R_1^k \times \sum_{i=0}^{\infty}(1-\pi)^i (i+1)^k =$$

$$= K_0(k,\pi)\mathbb{E}\,R_0^k + K(k,\pi)\mathbb{E}\,R_1^k,$$

где R_0 – это первый по времени момент регенерации процесса X_t', а $R_i \stackrel{\mathcal{D}}{=} R_1$, $i \in \mathbf{N}$ – это длина последующих (одинаково распределённых) периодов регенерации системы.

5.3. МЕТРИКА ПОЛНОЙ ВАРИАЦИИ

Далее, если $\mathcal{P}_t$ – распределение процесса X_t (в момент t), а $\mathcal{P}_t'$ – распределение процесса X_t', то

$$\|\mathcal{P}_t - \mathcal{P}_t'\|_{TV} \leqslant 2\frac{K_0(k,\pi)\mathbb{E}\,R_0^k + K(k,\pi)\mathbb{E}\,R_1^k}{t^k}.$$

Поскольку $\mathcal{P}_t' \Longrightarrow \mathcal{P}$ для всех X_0', получаем

$$\|\mathcal{P}_t - \mathcal{P}\|_{TV} \leqslant 2\frac{\int\limits_{\mathcal{X}} \Big(K_0(k,\pi)\mathbb{E}\,R_0^k + K(k,\pi)\mathbb{E}\,R_1^k\Big)\mathcal{P}(\,\mathrm{d}X_0')}{t^k} =$$

$$= 2\;\frac{K_0(k,\pi)\left(\int\limits_{\mathcal{X}} \mathbb{E}\,R_0^k \mathcal{P}(\,\mathrm{d}X_0')\right) + K(k,\pi)\mathbb{E}\,R_1^k}{t^k}.$$

Для интегрирования по мере (распределению) $\mathcal{P}$ надо его знать.

Однако нам неизвестно стационарное распределение процесса ни для процесса X_t, ни для хорошо изученной «стандартной» модели.

5.4. ОЦЕНКА СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ $\mathcal{P}$

Оценки распределения $\mathcal{P}$ снизу могут быть получены стандартными методами теории восстановления (см., например, [21]).

Поскольку для процесса восстановления с интенсивностью $v(s)$ и ф.р. $\Upsilon(s)$ времени восстановления $V_i \stackrel{\mathcal{D}}{=} V_1$ оценка стационарного распределения времени перескока $\widetilde{b}$ такова: $\mathbb{P}\left\{\widetilde{b} \leqslant a\right\} = \dfrac{\int\limits_a^\infty 1 - \Upsilon(u)\,\mathrm{d}u}{\int\limits_0^\infty x^2\,\mathrm{d}\Upsilon(x)}$, то при интегрировании по этой мере получим $\mathrm{d}\mathbb{P}\left\{\widetilde{b} \leqslant a\right\} = \dfrac{\Upsilon(a)}{\mathbb{E}(V_1)^2}$. Поэтому, учитывая условия раздела 2.2 и замечания 3, для интегрирования по стационарному распределению исследуемого процесса обслуживания по всему пространству $\mathcal{X}$ имеем:

$$\mathrm{d}\left(\mathcal{P}\{n_t = 0, x^{(0)} \leqslant a_0\}\right) \leqslant 1 \times \frac{\widetilde{\mathfrak{L}}(a_0)\,\mathrm{d}a_0}{\int\limits_0^\infty x^2\,\mathrm{d}\mathfrak{L}(x)};$$

$$\mathrm{d}\left(\mathcal{P}\left\{n_t = m > 0, x^{(0)} \leqslant a_0, x^{(1)} \leqslant a_1, x^{(2)} \leqslant a_2, \ldots \right.\right.$$
$$\left.\left.\ldots x^{(m)} \leqslant a_m,\right\}\right) \leqslant$$

$$\leqslant e^{-\rho}\frac{\rho^k}{k!} \times \frac{\widetilde{\mathfrak{L}}(a_0)\,\mathrm{d}a_0}{\int\limits_0^\infty x^2\,\mathrm{d}\mathfrak{L}(x)} \times \frac{Q(a_1)}{\int\limits_0^\infty x^2\,\mathrm{d}\Phi(x)} \times \frac{Q(a_2)}{\int\limits_0^\infty x^2\,\mathrm{d}\Phi(x)} \times \ldots$$
$$\ldots \times \frac{Q(a_m)}{\int\limits_0^\infty x^2\,\mathrm{d}\Phi(x)},$$

где $Q(x) = 1 - \exp\left(\int\limits_0^x -q(s)\,\mathrm{d}s\right)$.

Теперь, интегрируя $\int\limits_{\mathcal{X}} \mathbb{E}\, R_0^k \mathcal{P}(\,\mathrm{d}X_0')$ можно вычислить оценку константы $\mathbf{K}$.

5.5. ДРУГОЙ ПОДХОД К ВЫЧИСЛЕНИЮ ОЦЕНКИ ДЛЯ $\mathbf{K}$

Можно также рассмотреть вложенный процесс восстановления с периодами регенерации, равными моментам попадания регенерирующего процесса X_t в точки регенерации, т.е. моменты попадания в состояние $(1; 0; 0)$ – это моменты поступления в свободную СМО требования. Вложенный период восстановления такого регенерирующего процесса при выполнении условий раздела 2.2 позволяет оценивать степенную скорость сходимости с помощью неравенства Лордена. Для этого для исследуемого процесса восстановления надо найти оценки интенсивности восстановления по формулам (5)–(7), и эти оценки могут быть использованы по схемам, предложенным в [31, 32].

Благодарности

Автор благодарен $\boxed{\text{Л. Г. Афанасьевой}}$ за ценные обсуждения во время работы над статьёй. Также автор благодарит Э.Ю. Калимулину за большую помощь в подготовке статьи. Работа выполнена при поддержке РФФИ, проект № 20-01-00575 А.

Литература

1. АНИЧКИН С.А. *Склеивание процессов восстановления и его применение* // Проблемы устойчивости стохастических моделей. Труды семинара. – 1984. – М.: ВНИИСИ. – С. 4–24.
2. БОРИСОВ И.С. *Методы одного вероятностного пространства для марковских процессов* // Тр. Ин-та математики. – 1982. – Т. 1. – С. 4–18.
3. БОРОВКОВ А.А. *Обобщенные процессы восстановления*. – М.: Российская академия наук, 2020. – 453 с.

4. СЕВАСТЬЯНОВ Б.А. *Формулы Эрланга в телефонии при произвольном законе распределения длительности разговора* // Труды Третьего Всесоюзного математического съезда, Москва, июнь–июль 1956. – 1959. –Т. 4. – М.: Изд-во АН СССР. – С. 121–135.
5. СЕВАСТЬЯНОВ Б.А. *Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами* // Теория вероятн. и ее примен. – 1957. – Т. 2, вып. 1. – С. 106–116.
6. ШЕЛЕПОВА А.Д., САХАНЕНКО А.И. *Об асимптотике вероятности невыхода неоднородного обобщенного процесса восстановления за невозрастающую границу* // Сиб. электрон. матем. изв. – 2021. – Т. 18:2. – С. 1667—1688.
7. AFANASYEVA L.G., TKACHENKO A.V. *On the Convergence Rate for Queueing and Reliability Models Described by Regenerative Processes* // Journal of Mathematical Sciences. – October 2016. – Vol. 218, Issue 2. – P. 119–36.
8. ASMUSSEN S. *Applied probability and queues.* – Springer, New York, 2003.
9. CHANG J. T. *Inequalities for the Overshoot* // The Annals of Applied Probability. – 1994. – Vol. 4(4). – P. 1223.
10. DOEBLIN W. *Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états* // Rev. Math. de l'Union Interbalkanique. – 1938. – Vol.2. – P. 77–105.
11. ERLANG A. K. *Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges* // Elektroteknikeren. – 1917. – Vol. 13 – P. 5–13 (in Danish); Engl. transl.: P.O. Elect. Eng. Journal. – 1918. – Vol. 10. – P. 189–197; Reprinted as: WEB-based edition by permission from the Danish Acad. Techn. Sci. at http://oldwww.com.dtu.dk/teletraffic/Erlang.html.
12. FERREIRA M.A., ANDRADE M. *The $M|G|\infty$ queue busy period distribution exponentiality* // Journal of applied mathematics. – 2011. – Vol. 4, No. 3. – P. 249–260.

13. FORTET R. *Calcul des probabilités.* – CNRS, Paris, 1950.
14. GRIFFEATH D. *A maximal coupling for Markov chains* // und Zeitschrift für Wahrscheinlichkeitstheorie Verwandte Gebiete. – 1975. – Vol. 31, Iss. 2. – P. 95–106.
15. KALASHNIKOV V.V. *Mathematical Methods in Queuing Theory.* – Kluwer Academic Publishers, Amsterdam, 1994.
16. KALIMULINA E., ZVERKINA G. *On some generalization of Lorden's inequality for renewal processes* // arXiv.org. Cornell: Cornell university library. – 2019. – 1910.03381v1. – P. 1–5.
17. KATO K. *Coupling Lemma and Its Application to The Security Analysis of Quantum Key Distribution* // Tamagawa University Quantum ICT Research Institute Bulletin. – 2014. – Vol.4, No.1, – P. 23-30.
18. LORDEN G. *On Excess over the Boundary* // The Annals of Mathematical Statistics. – 1970. – Vol. 41(2). – P. 520–527.
19. PECHINKIN A.V. *On an invariant queueing system* // Math. Operationsforsch. Statist. Ser. Optim. – 1983. – Vol. 14(3). – P. 433–444.
20. PECHINKIN A.V., SOLOVYEV A.D., YASHKOV S.F. *A system with servicing discipline whereby the order of minimum remaining length is serviced first* // Eng. Cybern. – 1979. – Vol. 17(5). – P. 38–45.
21. SMITH W. L. *Renewal theory and its ramifications* // J. Roy. Statist. Soc. – 1958. – Ser. B, Vol. 20:2. – P. 243–302.
22. STADJE W. *The busy period of the queueing system* $M|G|\infty$ // Journal of Applied Probability. – 1985. – Vol. 22. – P. 697–704.
23. STOYAN D. *Qualitative Eigenschaften und Abschtzungen stochastischer Modelle.* – Berlin, 1977.
24. TAKÁCS L. *Introduction to the Theory of Queues.* – Oxford University Press, 1962.
25. VAN DOORN E.A., ZEIFMAN A.I. *On the speed of convergence to stationarity of the Erlang loss system* // Queueing Systems. – 2009. – Vol. 63. – P. 241–252.
26. VERETENNIKOV A.YU. *On rate of convergence to the stationary distribution in queueing systems with one serving*

device // Automation and Remote Control. – Vol. 74, Iss. 10. – P. 1620–1629. (2013).

27. VERETENNIKOV A.YU. *On the rate of mixing and convergence to a stationary distribution in Erlang-type systems in continuous time* // Problems Inf. Transmiss. –2010. – Vol. 46(4). – P. 382–389.
28. VERETENNIKOV A.YU. *The ergodicity of service systems with an infinite number of servomechanisms* // Математические заметки. – 1977. – Vol. 22(4). – P. 804–808.
29. VERETENNIKOV A., BUTKOVSKY O. A. *On asymptotics for Vaserstein coupling of Markov chains* // Stochastic Processes and their Applications. – 2013. – Vol. 123(9). – P. 3518–3541.
30. VERETENNIKOV A.YU., ZVERKINA G.A. *Simple proof of Dynkin's formula for single-server systems and polynomial convergence rates* // Markov Proc. Rel. Fileds. – 2014. – Vol. 20, Iss. 3. – P. 479–504; arXiv:1306.2359 [math.PR] (2013).
31. ZVERKINA G. *On strong bounds of rate of convergence for regenerative processes* // Communications in Computer and Information Science. – 2016. – Vol. 678. – P. 381–393.
32. ZVERKINA G. *Lorden's inequality and coupling method for backward renewal process* // Proc. of XX Int. Conf. on Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2017, Moscow), 2017. – P. 484–491.
33. ZVERKINA G. *On strong bounds of rate of convergence for regenerative processes* // Communications in Computer and Information Science – 2016. – Vol. 678. – P. 381–393.
34. ZVERKINA G. *About some extended Erlang-Sevast'yanov queueing system and its convergence rate* (English and Russian versions) // https:\\arxiv.org/abs/1805.04915; Фундаментальная и прикладная математика. – 2018. – No. 22, Iss. 3. – P. 57–82.
35. ZVERKINA G., KALIMULINA E. *On generalized intensity function and its application to the backward renewal time*

estimation for renewal processes // Proc. of the 5th Int. Conf. on Stochastic Methods (ICSM-5, 2020). – 2020. – M.: Изд-во РУДН. – P. 306–310.

LORDEN'S INEQUALITY AND THE RATE OF CONVERGENCE OF THE DISTRIBUTION OF ONE GENERALIZED ERLANG – SEVAST'YANOV QUEUING SYSTEM

Galina Zverkina, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc., assistant professor (zverkina@gmail.com).

Abstract: It is more important to estimate the rate of convergence to a stationary distribution rather than only to prove the existence one in many applied problems of reliability and queuing theory. This can be done via standard methods, but only under assumptions about an exponential distribution of service time, independent intervals between recovery times, etc. Results for such simplest cases are well-known. Rejection of these assumptions results to rather complex stochastic processes that cannot be studied using standard algorithms. A more sophisticated approach is needed for such processes. That requires generalizations and proofs of some classical results for a more general case. One of them is the generalized Lorden's inequality proved in this paper. We propose the generalized version of this inequality for the case of dependent and arbitrarily distributed intervals between recovery times. This generalization allows to find upper bounds for the rate of convergence for a wide class of complicated processes arising in the theory of reliability. The rate of convergence for a two-component process has been obtained via the generalized Lorden's inequality in this paper.

Keywords: regenerative Markov processes, coupling method, total variation metric, generalized Erlang-Sevastyanov system, rate of convergence.

УДК 519.21
ББК 22.171
DOI: 10.25728/ubs.2023.102.2

Статья представлена к публикации
членом редакционной коллегии В.М. Вишневским.

Поступила в редакцию 14.11.2022
Дата опубликования 31.03.2023.