ФГБОУ ВО
Московский Государственный Университет
им. М.В.Ломоносова
Механико-Математический Факультет

На правах рукописи

УДК

Айбатов Серик Жагалбаевич

МОДЕЛИ ТЕОРИИ ОЧЕРЕДЕЙ С ПРЕРЫВАНИЕМ ОБСЛУЖИВАНИЯ

01.01.05 — теория вероятностей и математическая статистика

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель —
д.ф.-м.н. Л.Г. Афанасьева

Москва, 2016

Оглавление

Введение

Диссертация подготовлена на кафедре теории вероятностей механико-математического факультета Московского государственного университета им. М. В. Ломоносова и затрагивает ряд вопросов, относящихся к теории массового обслуживания (теории очередей) и теории случайных процессов.

Актуальность и история вопроса

Диссертация посвящена исследованию систем массового обслуживания, в которых прибор время от времени недоступен для обслуживания, т.е. в случайные моменты происходит прерывание обслуживания. Будем называть такие моменты - поломкой прибора, а интервалы, когда прибор не может обслуживать требования - периодами восстановления или ремонта.

Основное внимание уделяется отысканию условий эргодичности систем с ненадежным прибором, нахождению операционных характеристик в стационарном режиме и оценке вероятности образования большой очереди.

Системы с прерываниями обслуживания изучаются довольно давно. По-видимому, первой работой по данной тематике является статья [72](White, Christie, 1958) в которой рассмотрена система $M/M/1/\infty$ с ненадежным прибором, где времена ремонта и рабочего состояния прибора имеют экспоненциальные распределения с разными параметрами. Одной из наиболее частых причин прерывания обслуживания является поступление в систему требования с более высоким приоритетом. Статья [62](Miller, 1960) посвящена такому виду прерывания. Отдельно стоит отметить работу [47](Gaver, 1962), где рассмотрена система $M/G/1/\infty$, прибор в которой может ломаться только во время обслуживания требования, время рабочего состояния распределено экспоненциально, а время ремонта имеет произвольное распределение. В этой работе было введено понятие "полного времени обслуживания"(completion time), которое позволяет свести анализ системы с ненадежным прибором к классической системе $M/G/1/\infty$. Также там было рассмотрено несколько вариантов обслуживания требования после прерывания. Практически аналогичная модель была рассмотрена в книге [18](Бочаров, Печинкин, 1995) с той разницей, что поломка прибора теперь может происходить в любое время. О работах [47] и [18] более подробно написано в главе 2. Анализу подобных систем посвящено очень много работ, например, [41](Doshi, 1986), [52](Keilson, 1962). В обзорной статье [58](Krishnamoorthy et. al. 2012) дано подробное описание имеющихся на данный момент в этом направлении результатов. Одной из последних работ является статья [28](Афанасьева, Баштова, 2014), где рассмотрена система $Reg/G/1/\infty$ с ненадежным прибором, времена восстановления и рабочего состояния которого независимы и произвольно распределены.

Определение условий эргодичности процессов, описывающих функционирование систем, является одной из первых задач, которые приходится решать при анализе систем обслуживания. Эти условия представляют собой соотношения между параметрами модели, при которых не образуется бесконечно больших очередей. Доказательства соответствующих теорем приводят к анализу достаточно сложных случайных процессов, вообще говоря, не марковских. Если же удается построить цепь Маркова, описывающую функционирование системы, то доказательства предельных теорем основываются на результатах для марковских процессов. Одними из первых работ в данном направлении являются статьи [20] (Kendall, 1959) и [45] (Foster, 1953), в которых найдены достаточные условия существования стационарного распределения у цепей Маркова, связанных с очередью в системе. Монография [15] (Боровков, 1999) посвящена анализу свойств эргодичности и устойчивости широкого класса случайных процессов (цепей Маркова, стохастически рекурсивных последовательностей и рекурсивных цепей и др.). Исследование систем обслуживания часто сводится к изучению марковских процессов с помощью введения дополнительной переменной. Этот метод использован, например, в [24] (Севастьянов, 1957) для исследования систем с отказами при произвольном распределении времени обслуживания, а также в [21] (Коваленко, 1961) для систем с ограничениями. Другой метод доказательства эргодических теорем состоит в построении процессов, стохастически монотонных по времени. В этом случае из монотонности следует существование предела последовательности функций распределения. Условия, при которых этот предел задает распределение вероятностей, могут быть получены с помощью метода, предложенного в [61] (Loynes, 1962), (см., например, [1] (Афанасьева, 1965), [3] (Афанасьева, Мартынов, 1969)).

Естественным и актуальным направлением развития теории является исследование моделей с входными потоками более общего вида. В настоящей работе представлены результаты для систем с регенерирующим входным потоком и ненадежным прибором. Причем интервалы рабочего состояния прибора и ремонта определяются некоторым регенерирующим процессом, т.е. в нашей системе данные интервалы могут быть зависимыми, что является обобщением существующих моделей с ненадежным прибором. Для такой системы найдено условие эргодичности процесса виртуального времени ожидания.

Класс регенерирующих потоков обладает рядом полезных свойств.

Во-первых, регенерирующими являются большинство потоков, обычно используемых в теории массового обслуживания в качестве входных. Среди них дважды стохастический пуассоновский поток со случайной интенсивностью, являющейся регенерирующим процессом (J. Grandell, 1976 [49]), марковски-модулированный поток (S. Asmussen, 1991 [31]), марковский поток поступлений (V. Klimenok et al., 2005 [55]), полумарковский поток (S. Asmussen, 1999 [32], Л. Г. Афанасьева, Е. Е. Баштова, Е. В. Булинская, 2009 [29]).

Во-вторых, при довольно общих условиях свойство регенерации сохраняется при прохождении через систему обслуживания. Это позволяет исследовать последовательно соединённые системы обслуживания и иерархические сети, опираясь на результаты, касающиеся отдельных узлов, подобно тому, как это сделано в [4](Афанасьева, 1987).

Если известно, что система эргодична и со временем очередь не растет до бесконечности, то возникает важная проблема, связанная с оценкой вероятности образования сколь угодно большой очереди. То есть, если $\Psi(x)$ — предельная функция распределения остаточного времени ожидания, то нас интересует асимптотика функции $1 - \Psi(x)$ при $x \to \infty$. Данному вопросу посвящено большое количество работ. В статье [27](Abate, Whitt, 1997) данная задача решена для системы $M/G/1/\infty$ с приоритетами, в случае "тяжелых хвостов"и "легких". Вариант легких хвостов и система $GI/G/1/\infty$ был рассмотрен в работах [48](Glynn, Whitt, 1994), [46](Ganesh et. al. 2004), в статье [7](Афанасьева, Баштова, 2015) этот результат обобщен на систему $Reg/G/1/\infty$. В диссертации мы будем рассматривать только тяжелые хвосты. Как отмечено в [43](Фосс и др. 2007), распределения с тяжелыми хвостами играют существенную роль при моделировании коммуникационных сетей. В силу различия в запросах от разных групп потребителей, часть трафика касается малых объемов работы, но есть запросы весьма большого объема и это приводит к распределениям с тяжелыми хвостами (см. также [59](Leland et. al. 1994), [66](Resnick, 1997), [73](Willinger et. al. 1995)).

Хорошо известно [60](Lindley, 1952), что время ожидания n-го требования ($n = 1, 2, \dots$) в одноканальной системе определяется рекуррентным соотношением (при $w_1 = 0$)

$$w_{n+1} = \max(0, w_n + \eta_n - \zeta_n) = \max(0, w_n + u_n),$$

где η_n — время обслуживания n-го требования, ζ_n — интервал между моментами поступлений $(n+1)$-го и n-го требований, $u_n = \eta_n - \zeta_n$. Если $\{u_n\}_{n=1}^{\infty}$

стационарная последовательность и $w_1 = 0$, то по распределению

$$w_n \stackrel{d}{=} \max_{0 \leqslant k \leqslant n-1} S_k,$$

где $S_k = \sum_{j=1}^{k} u_j$ и $S_0 = 0$. Отсюда следует, что существует

$$F(x) = \lim_{n \to \infty} \mathsf{P}(w_n \leqslant x) = \mathsf{P}\left(\sup_{n \geqslant 0} S_n \leqslant x\right) \tag{1}$$

и $F(x)$ — функция распределения, когда $\mathsf{E}u_n < 0$ [61](Loynes, 1962).

Первые результаты в оценке вероятностей больших уклонений были получены для систем с управляющей последовательностью $\{u_n\}_{n=1}^{\infty}$, состоящей из независимых одинаково распределенных случайных величин на основе соотношения (1). Для регулярно меняющихся на $+\infty$ функций $G(x) = \int_x^{\infty} \mathsf{P}(u_n > y)dy$ асимптотика

$$\mathsf{P}\left(\sup_{n \geqslant 0} S_n > x\right) \sim \frac{G(x)}{-\mathsf{E}u_n} \quad (x \to \infty) \tag{2}$$

была установлена в работе [38](Cohen, 1969), а затем обобщена на более широкий класс функций (надстепенных) в статье [12](Боровков, 1970, см., также [11], [13], [16], [37]). Для субэкспоненциальных функций $G(x)$ асимптотика (2) доказана в [64](Pakes, 1975),[71](Veraverbeke, 1977). Дальнейшее развитие связано, с так называемыми, модулированными случайными блужданиями, для которых стационарная последовательность $\{u_n\}_{n=1}^{\infty}$ состоит, вообще говоря, из зависимых случайных величин. Предполагается, что распределение приращения случайного блуждания $u_n = S_n - S_{n-1}$ определяется значением некоторого случайного процесса X_n. Отметим несколько статей на эту тему. Арндт ([35],1980) рассмотрел приращения с регулярно меняющимися хвостами, модулированные цепью Маркова с конечным множеством состояний. Алсмейр и Сбигнев ([34],1999), а также Йеленкович и Лазар ([51],1998), исследовали распределение верхней грани случайного блуждания, модулированного цепью Маркова с конечным множеством состояний, но в предположении, что приращения u_n имеют субэкспоненциальный интегральный хвост. В работах [32](Asmussen, 1999) и [33](Asmussen et. al.,1999) рассмотрены случайные блуждания $Y(t)$ регенерирующей структуры. В предположении, что распределение верхней грани приращения $Y(t)$ на периоде регенерации асимптотически (при $x \to \infty$) эквивалентно распределению приращения за то же время, для случая субэкспоненциального интегрального хвоста получена асимптотика распределения верхней грани. Результаты применены к анализу системы обслуживания с полумарковским входящим потоком, для которой соответствующее случайное блуждание может рассматриваться как модулированное.

В качестве модулирующей среды выступает полумарковский процесс. Установлены условия субэкспоненциальной асимптотики распределения процесса виртуального времени ожидания $W(t)$ в стационарном режиме. Ряд результатов для модулированных случайных блужданий получен в работе [43](Foss et. al., 2007). Доказаны достаточно общие условия субэкспоненциальности распределения верхней грани. Рассмотрена ситуация, когда модулирующий процесс имеет регенерирующую структуру.

В диссертации найдена асимптотика больших уклонений процесса виртуального времени ожидания в стационарном режиме для системы с регенерирующим входящим потоком, в предположении, что суммарная работа, поступившая в систему за период регенерации, имеет субэкспоненциальный интегральный хвост. Данный результат распространен на вложенные процессы w_n и $W_n = W(\theta_n)$, где θ_n — n-я точка регенерации входного потока. Эта задача также решена в случае, когда прибор ненадежен.

Цели работы

Целью настоящей диссертационной работы являются:

- определение условий стабильности системы $Reg/G/1/\infty$ с ненадежным прибором, когда функционирование прибора определяется регенерирующим процессом, не зависящим от входящего потока;
- оценка вероятностей больших уклонений для функции распределения виртуального времени ожидания в стационарном режиме для системы $Reg/G/1/\infty$ с надежным и ненадежным прибором в случае "тяжелых" хвостов;
- нахождение явного распределения виртуального времени ожидания и количества требований в системе в стационарном режиме для системы $M/G/1/\infty$ с ненадежным прибором и наличием приоритетных требований.

Научная новизна работы

Все результаты работы являются новыми. В диссертации получены следующие основные результаты.

1. Для системы обслуживания $Reg/G/1/\infty$ с ненадежным прибором, в которой процесс, описывающий моменты поломок и восстановления прибора, является регенерирующим, не зависящим от входного потока, установлены необходимые и достаточные условия эргодичности.

2. Для системы обслуживания $M/G/1/\infty$ с ненадежным прибором и приоритетными требованиями получены условия эргодичности, найдены предельные распределения процессов виртуального времени ожидания и числа требований в системе в терминах преобразований Лапласа-Стилтьеса, приведены выражения для важных операционных характеристик. Анализ проведен для двух дисциплин обслуживания требований после прерывания.

3. Найдена асимптотика вероятностей больших уклонений процесса виртуального времени ожидания для системы $Reg/G/1/\infty$ в предположении, что суммарное время обслуживания требований, поступивших в течение периода регенерации, имеет "тяжелый хвост". Аналогичный результат получен для системы $Reg/G/1/\infty$ с ненадежным прибором.

Методы исследования

В диссертации используются различные методы и результаты теории вероятностей и теории случайных процессов: предельные теоремы для стационарных процессов, теорема восстановления Блекуэлла, узловая теорема Смита, метод склеивания (coupling), теоремы о регенерирующих процессах, метод полного времени обслуживания.

Практическая и теоретическая ценность

Диссертация носит теоретический характер. Её результаты могут найти применение в теории очередей, теории надежности, а также использоваться при исследовании телекоммуникационных систем.

Апробация диссертации

По теме диссертации были сделаны доклады на следующих семинарах механико-математического факультета МГУ им. М. В. Ломоносова:

- Большом семинаре кафедры теории вероятностей под руководством действительного члена РАН, профессора А. Н. Ширяева (2016 г.),

- Семинаре «Исследование асимптотического поведения и устойчивости стохастических моделей» под руководством проф. Л.Г. Афанасьевой, проф. Е.В. Булинской, доц. Е.Б. Яровой (2013–2016 гг., неоднократно).

Результаты диссертации докладывались на:

- Международной конференции “XXXII International Seminar on Stability Problems for Stochastic Models” в Норвегии (Тронхейм, 2014 г.);

- Международной конференции “16th Applied Stochastic Models and Data Analysis International Conference (ASMDA2015)” в Греции (Пирей, 2015 г.);
- Международной конференции студентов, аспирантов и молодых учёных «Ломоносов» в МГУ (Москва, 2013-2016 гг.);
- Международной конференции по стохастическим методам (г. Новороссийск, пос. Абрау-Дюрсо, 2016 г.);
- Международной конференции по исследованию операций (ORM2016) в России (Москва, 2016г.).

Публикации

Основные результаты диссертации содержатся в работах [74-85], представленных в конце списка литературы. Среди них четыре статьи в журналах из перечня ВАК.

Содержание диссертации

Во **введении** приводится краткий обзор исследований, посвящённых системам с ненадежными приборами, вопросам эргодичности различных систем и оценке вероятностей больших уклонений процесса виртуального времени ожидания. Историческая справка подкрепляется соответствующими ссылками на научные работы. Кроме того, во введении объясняется актуальность темы и научная новизна предпринятого автором исследования. Далее нумерация утверждений и условий совпадает с их нумерацией в соответствующих главах.

Первая глава посвящена установлению условий, при которых одноканальная система обслуживания с ненадежным прибором и регенерирующим входящим потоком является стабильной.

В разделе 1.1 даются основные определения, такие как регенерирующий процесс, регенерирующий поток, стохастическая ограниченность и эргодичность процесса. Приводятся используемые далее результаты, относящиеся к регенерирующим процессам и процессам восстановления (узловая теорема Смита, теорема Блекуэлла, основная теорема восстановления и др.). В этом разделе определяется следующий класс функций распределений.

Условие 1. Пусть τ случайная величина, тогда существует $n \geqslant 1$ и неотрицательная функция $f(x)$ такие, что $\int\limits_{\mathbb{R}} f(x)dx > 0$ и

$$\mathsf{P}(\hat{\tau}_1 + \cdots + \hat{\tau}_n \in A) \geqslant \int_A f(x)dx, \quad A \in \mathcal{B},$$

где $\hat{\tau}_1, \ldots, \hat{\tau}_n$ независимые одинаково распределенные случайные величины, распределенные как τ. В зарубежной литературе такие распределения называются — *spread out distributions* (см. [30, Глава 7](Asmussen, 1987)).

В разделе 1.2 показано, что класс регенерирующих процессов довольно широк и он включает в себя многие входные потоки, используемые в теории массового обслуживания. В качестве примеров таких потоков упоминаются дважды стохастический пуассоновский поток (ДСПП) с регенерирующей интенсивностью, поток с интенсивностью случайной амплитуды, поток со случайными периодами, марковски-модулированный поток (ММП), поток Льюиса, марковский поток поступлений (МПП), полумарковский поток (ПМП).

Далее,в разделе 1.3 дается описание изучаемой модели. Рассматривается одноканальная система массового обслуживания с неограниченной очередью и регенерирующим входящим потоком $X(t)$, определенным на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ с фильтрацией $\{\mathscr{F}_t, t \geqslant 0\}$, к которой он адаптирован. Марковские моменты $\{\theta_i^{(1)},\, i \geqslant 1\}$ $(\mathsf{P}(\theta_0^{(1)} < \infty) = 1)$ относительно фильтрации $\{\mathscr{F}_t, t \geqslant 0\}$ — точки регенерации потока, $\tau_i^{(1)} = \theta_i^{(1)} - \theta_{i-1}^{(1)}$ — длина i-го периода регенерации, $\gamma_i^{(1)} = X(\theta_i^{(1)}) - X(\theta_{i-1}^{(1)})$. Предполагается, что $\mathsf{E}\tau_2^{(1)} < \infty$, $\mathsf{E}\gamma_2^{(1)} < \infty$, тогда интенсивность потока $\lambda_x = \lim\limits_{t\to\infty} \frac{X(t)}{t} = \frac{\mathsf{E}\gamma_2^{(1)}}{\mathsf{E}\tau_2^{(1)}}$.

Здесь процесс $X(t)$ — это объем работы, который поступил в систему за время $[0, t]$.Заметим, что мы не накладываем никаких условий на отдельные времена обслуживания, они могут быть зависимыми в рамках одного периода регенерации.

Пусть $Y(t)$ — количество работы, которое может выполнить система за время $(0, t]$, если бы в системе всегда были требования. Считаем, что процессы $X(t)$ и $Y(t)$ не зависят друг от друга. В классической модели, когда прибор надежен и работает с единичной скоростью $Y(t) = t$. Здесь мы предполагаем, что

$$Y(t) = \int_0^t e(s)ds,$$

где $e(t, w)$ неотрицательный случайный процесс, все траектории которого интегрируемы по Лебегу в $(0, t]$. Кроме того, мы считаем, что процесс $\{e(t),\, t \geqslant 0\}$ — регенерирующий с точками регенерации $\{\theta_i^{(2)}\}_{i=1}^{\infty}$ $(\mathsf{P}(\theta_0^{(2)} < \infty) = 1)$ и случайные величины $\tau_i^{(2)} = \theta_i^{(2)} - \theta_{i-1}^{(2)}$, $\zeta_j = Y(\theta_j) - Y(\theta_{j-1}) = \int\limits_{\theta_{j-1}}^{\theta_j} e(s)ds$ имеют конечные математические ожидания

$$\mathsf{E}\tau_2^{(2)} < \infty, \quad \mathsf{E}\zeta_2 < \infty.$$

Тогда $Y(t)$ будет регенерирующим потоком и его интенсивность $\lambda_y = \frac{\mathsf{E}\zeta_2}{\mathsf{E}\tau_2^{(2)}}$, а *коэффициент загрузки системы*

$$\rho = \frac{\lambda_x}{\lambda_y} = \frac{\mathsf{E}\gamma_2^{(1)}}{\mathsf{E}\zeta_2}\frac{\mathsf{E}\tau_2^{(2)}}{\mathsf{E}\tau_2^{(1)}}.$$

В модели, когда прибор может находиться в рабочем или нерабочем (он ремонтируется) состоянии, процесс

$$e(t) = \begin{cases} 0, & \text{прибор сломан} \\ 1, & \text{иначе.} \end{cases}$$

В данной модели предполагается, что при поломке прибора происходит немедленное прерывание обслуживания требования. Прерванное обслуживание продолжается после восстановления прибора, причем время обслуживания требования после восстановления прибора совпадает с остаточным временем обслуживания этого требования в момент поломки.

Введем процесс виртуального времени ожидания $W(t)$, представляющий собой суммарное остаточное время обслуживания требований, находящихся в системе в момент t. Как известно (см., например, [68](Такач, 1967),[13](Боровков, 1972)), для данной модели будет справедливо следующее равенство

$$W(t) = \max[W(0) + Z(t), \sup_{0\leqslant s\leqslant t}(Z(t) - Z(s))], \tag{3}$$

где $Z(t) = X(t) - Y(t)$. Приведем условие, при котором процесс $Z(t)$ при $t \to \infty$ стремится по распределению к процессу $\tilde{Z}(t)$ со стационарными приращениями.

Следуя общепринятой терминологии в теории восстановления, мы рассматриваем 2 случая:

- дискретный, когда все случайные величины, определяющие модель, имеют решетчатые распределения с одним и тем же шагом;
- непрерывный случай, когда эти распределения нерешетчатые.

Условие 2. В непрерывном случае: распределения случайных величин $\tau_2^{(1)}$ и $\tau_2^{(2)}$ удовлетворяют условию 1. В дискретном случае: распределения случайных величин $\tau_2^{(1)}$ и $\tau_2^{(2)}$ апериодичны.

Распределение случайной величины τ *апериодично*, если наибольший общий делитель $\{j : \mathsf{P}(\tau = j) > 0\}$ равен единице.

Лемма 1. *Если существует предельное распределение* $\pi(x) = \lim\limits_{t\to\infty} \mathsf{P}(e(t) \leqslant x)$, *то интенсивность потока* $Y(t)$ *равна* $\lambda_y = \int\limits_0^\infty x d\pi(x)$. *Если процесс* $e(t)$

принимает только два значения $\{0,1\}$ *и существует* $\lim\limits_{t\to\infty} \mathsf{P}(e(t)=1)=\pi$, *то* $\lambda_y=\pi$.

Следующее условие позволяет доказать наличие общих точек регенерации для потоков $X(t)$ и $Y(t)$.

Условие 3. Для n-го ($n \geqslant 1$) периода регенерации потока $Y(t)$ имеет место представление:

$$\tau_n^{(2)} = \upsilon_n^{(1)} + \upsilon_n^{(2)},$$

где $\upsilon_n^{(1)}$ и $\upsilon_n^{(2)}$ независимы, $\mathsf{P}(\upsilon_n^{(1)} > x) = e^{-\delta x}$, $\delta \in (0,\infty)$ и $Y(\theta_{n-1}^{(2)}+\upsilon_n^1) = Y(\theta_{n-1}^{(2)})$ с вероятностью 1.

Лемма 2 (О синхронизации). *Пусть выполнено условие 2 в дискретном случае и условие 3 в непрерывном случае. Определим в дискретном случае*

$$T_0 = \min\left\{\theta_j^{(1)} : \bigcup_{l=0}^{\infty}(\theta_j^{(1)} = \theta_l^{(2)})\right\},$$

$$T_n = \min\left\{\theta_j^{(1)} > T_{n-1} : \bigcup_{l=0}^{\infty}(\theta_j^{(1)} = \theta_l^{(2)})\right\},$$

соответственно в непрерывном случае

$$T_0 = \min\left\{\theta_j^{(1)} : \theta_j^{(1)} \in \bigcup_{l=1}^{\infty}(\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + \upsilon_l^{(1)})\right\},$$

$$T_n = \min\left\{\theta_j^{(1)} > T_{n-1} : \theta_j^{(1)} \in \bigcup_{l=1}^{\infty}(\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + \upsilon_l^{(1)})\right\}.$$

Тогда $\{T_n\}_{n=0}^{\infty}$ *— общие моменты регенерации для процессов* $X(t)$, $Y(t)$ *и*

в дискретном случае $\quad \mathsf{E}\tau_2 = \mathsf{E}\tau_2^{(1)}\mathsf{E}\tau_2^{(2)} < \infty,$

в непрерывном случае $\quad \mathsf{E}\tau_2 = \delta\mathsf{E}\tau_2^{(1)}\mathsf{E}\tau_2^{(2)} < \infty,$

где $\tau_n = T_n - T_{n-1}$, $n \geqslant 1$.

В разделе 1.4 приводится основная теорема главы, в которой показаны необходимые и достаточные условия, при которых процесс $W(t)$ будет эргодичным.

Теорема 1. *Пусть* $W(t)$ *случайный процесс, определяемый по формуле* (3), *тогда*

- *Если* $\rho > 1$, *то* $W(t) \xrightarrow[t\to\infty]{} \infty$ *с вероятностью 1.*
- *Если* $\rho < 1$ *и выполнено условие 2, то существует* $\lim\limits_{t\to\infty}\mathsf{P}(W(t) \leqslant x) = \Psi(x)$ *и* $\Psi(x)$ *является функцией распределения, т.е. процесс* $W(t)$ *эргодичен.*

- *Если $\rho = 1$ и выполнено условие 3, то $W(t) \xrightarrow[t\to\infty]{} \infty$ по вероятности.*

Далее, этот результат переносится на систему обслуживания $Reg/G/1/\infty$ с ненадежным прибором. Для этого введем входящий поток $A(t)$ — количество требований, поступивших в систему за время $(0,t]$. Поток $A(t)$ — регенерирующий с точками регенерации $\{\theta_n^{(1)}\}_{n=0}^{\infty}$. Пусть $\xi_n = A(\theta_n^{(1)}) - A(\theta_{n-1}^{(1)})$ и $\mathsf{E}\xi_n < \infty$, тогда для интенсивности потока $A(t)$ справедливо равенство $\lambda = \frac{\mathsf{E}\xi_n}{\mathsf{E}\tau_n^{(1)}}$.

Времена обслуживания описываются последовательностью $\{\eta_n\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин, $\mathsf{E}\eta_n = b < \infty$. Причем эта последовательность не зависит от процесса $A(t)$. Тогда поток $X(t)$ определяется следующей формулой

$$X(t) = \sum_{k=1}^{A(t)} \eta_k.$$

Он тоже будет регенерирующим с точками регенерации $\{\theta_n^{(1)}\}_{n=0}^{\infty}$.

Следствие 1. *Теорема 1 верна для системы $Reg/G/1/\infty$ с ненадежным прибором и коэффициент загрузки принимает следующий вид*

$$\rho = \frac{\lambda b}{\pi}.$$

В разделе 1.5 даются примеры приложения следствия 1.

Вторая глава посвящена исследованию моделей типа $M/G/1/\infty$ с ненадежным прибором и приоритетным обслуживанием. Анализ этих моделей проводится параллельно для двух дисциплин обслуживания требований после прерывания:

D1 (*Preemptive-resume interruptions*). Прерывание обслуживания интерпретируется как поломка прибора и, соответственно, имеет немедленный эффект. Прерванное обслуживание продолжается после восстановления прибора, причем время обслуживания требования после восстановления прибора совпадает с остаточным временем обслуживания этого требования в момент поломки.

D2 (*Preemptive-repeat-different interruptions*). Обслуживание требования прерывается немедленно, а время обслуживания требования после восстановления прибора имеет то же распределение, что и начальное время обслуживания, и не зависит от него.

В разделе 2.1 даются далее используемые результаты из [18](Бочаров, Печинкин, 1995) и [47](Gaver, 1962).

В [18] рассматривается система массового обслуживания $M/G/1/\infty$ с ненадежным прибором. Вероятность выхода прибора из строя на интервале времени $(t, t+\Delta)$ зависит только от состояния системы в момент t и, если система в момент t свободна, то она равна $\mu^*\Delta + o(\Delta)$, а если на приборе находится требование, то $\mu\Delta + o(\Delta)$. Таким образом, параметры μ^* и μ — интенсивности отказов в свободном и занятом состояниях соответственно.

Если в момент отказа прибора система свободна, то прибор ремонтируется случайное время, имеющее функцию распределения $G^*(x)$ с преобразованием Лапласа-Стилтьеса (ПЛС) $\zeta^*(s)$ и средним g^*. При этом первое требование, поступившее в свободную систему в момент ремонта прибора, становится на прибор, но его обслуживание начинается только после окончания ремонта. Остальные поступающие требования скапливаются в очереди.

Если же в момент отказа прибора на нем находится требование, то обслуживание прекращается, а прибор ремонтируется случайное время с функцией распределения $G(x)$, ПЛС $\zeta(s)$ и средним g. В течение этого времени требование продолжает находиться на приборе, причем после восстановления прибора требование дообслуживается остаточное время.

Время обслуживания требования имеет функцию распределения $B(x)$, математическое ожидание b и ПЛС $\beta(s)$.

При помощи результатов из [47], мы находим $\alpha(s)$ и $\alpha^*(s)$ — ПЛС времени пребывания требования на приборе (полное время обслуживания), пришедшего в занятую систему и свободную соответственно для обеих дисциплин обслуживания после прерывания:

$$\mathsf{D1} \qquad \alpha(s) = \beta(s + \mu - \mu\zeta(s)), \tag{4}$$

$$\mathsf{D2} \qquad \alpha(s) = \frac{\beta(s+\mu)}{1 - \zeta(x)\frac{\mu}{s+\mu}(1 - \beta(s+\mu))}. \tag{5}$$

Здесь λ — интенсивность поступления требований в систему. Формулу для $\alpha^*(s)$ можно найти в [18], она имеет следующий вид

$$\alpha^*(s) = \frac{\lambda}{\lambda + \mu^* - \mu^*\zeta^*(\lambda)} \left[1 + \mu^* \frac{\zeta^*(s) - \zeta^*(\lambda)}{\lambda - s}\right] \alpha(s). \tag{6}$$

Введем процессы $q(t)$ и $W(t)$ — число требований в системе и виртуальное время ожидания в момент t соответственно. Пусть $\rho = \lambda a < 1$, где $a = -\alpha'(0)$. Тогда эти процессы имеют предельные стационарные распределения

$$p_i = \lim_{t\to\infty} \mathsf{P}(q(t) = i), \quad \Psi(x) = \lim_{t\to\infty} \mathsf{P}(W(t) \leqslant x).$$

При этом, если $P(z)=\sum_{i=0}^{\infty}p_iz^i$ и $w(s)=\int_0^{\infty}e^{-sx}d\Phi(x)$, то

$$P(z)=\frac{\alpha(\lambda-\lambda z)-z\alpha^*(\lambda-\lambda z)}{\alpha(\lambda-\lambda z)-z}p_0,$$
$$w(s)=\frac{s+\mu^*-\mu^*\zeta^*(s)}{s-\lambda+\lambda\alpha(s)}\frac{1-\rho}{1+\mu^*g^*}, \tag{7}$$

где $p_0=\frac{1-\rho}{1-\rho+\lambda a^*}$, $a^*=-(\alpha^*)'(0)$.

В разделе 2.2 мы находим асимптотику функции $1-\Psi(x)$ при $x\to\infty$, предполагая, что распределения времени обслуживания и времен ремонта имеют регулярно меняющиеся хвосты, т.е.

Условие 5. Пусть l,m и m^* целые числа, большие 1, и при $x\to\infty$

$$1-B(x)\sim\frac{(-1)^l}{\Gamma(1-p)}x^{-p}L_1(x),\quad l<p<l+1,$$
$$1-G(x)\sim\frac{(-1)^m}{\Gamma(1-r)}x^{-r}L_2(x),\quad m<r<m+1,$$
$$1-G^*(x)\sim\frac{(-1)^{m^*}}{\Gamma(1-r^*)}x^{-r^*}L_2^*(x),\quad m^*<r^*<m^*+1,$$

где $L_1(x),L_2(x)$ и $L_2^*(x)$ — медленно меняющиеся функции.

Используя формулу 7 и теорему из [36](Bingham,Doney, 1974), мы доказываем теорему.

Теорема 2. *Пусть требования после прерывания обслуживаются по дисциплине* D1. *Если выполнено условие 5, то при* $x\to\infty$

$$1-\Psi(x)\sim\frac{(-1)^{\tilde n-1}}{\Gamma(2-\tilde q)}x^{-\tilde q+1}L_4(x),$$

где $\tilde n=\min(l,m,m^*)$, $\tilde q=\min(p,r,r^*)$ *и*

$$L_3(x):=1(p\leqslant r)(1+\mu g_1)^pL_1(x)+1(p\geqslant r)\mu b_1L_2(x),$$
$$L_4(x):=1(q\leqslant r^*)\frac{\lambda}{1-\rho}L_3(x)+1(q\geqslant r^*)\frac{\mu^*}{1+\mu^*g_1^*}L_2^*(x).$$

Аналогичный результат для системы с дисциплиной D2 приведен в следующей главе, причем для более общего входящего потока и сильно субэкспоненциальных функций распределения времени обслуживания.

Далее мы рассматриваем две модели (M_1 и M_2), в которых присутствуют два вида прерываний обслуживания (поломка прибора, поступление приоритетного требования). В обеих моделях мы предполагаем, что входящие потоки являются пуассоновскими соответственно с параметрами λ_1 и λ_2. Времена обслуживания требований i-го типа ($i=1,2$) образуют последовательность

$\{\eta_n^i\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин с функцией распределения $B_i(x)$, средним b_i и ПЛС $\beta_i(s)$.

В модели M_1 считаем, что прибор выходит из строя и восстанавливаются независимо от того, обслуживает он требование или нет. При этом функционирование прибора определяется двумя независимыми последовательностями, каждая из которых состоит из независимых одинаково распределенных случайных величин. При этом время рабочего состояния распределено экспоненциально с параметром ν, а время восстановления имеет функцию распределения $D(x)$, среднее d и ПЛС $\delta(s)$.

В модели M_2 предполагается, что прибор состоит из двух блоков, первый обслуживает требования первого типа, а второй — требования второго. При этом в любой момент времени может работать только один блок. Каждый блок может выйти из строя во время обслуживания требования. Время рабочего состояния i-го блока ($i = 1, 2$) имеет экспоненциальное распределение с параметром μ_i, а время восстановления распределено по закону $G_i(x), \zeta_i(s) = \int_0^\infty e^{-sx} dG_i(x)$ и имеет математическое ожидание g_i.

В разделах 2.4 и 2.6 мы показываем как M_1 и M_2 сводятся к модели, описанной в разделе 2.1. Благодаря этому мы находим стационарные распределения процессов $q(t)$ и $W(t)$ для этих моделей, а также получаем асимптотику вероятностей больших уклонений функции $1 - \Psi(x)$ при $x \to \infty$. Данные результаты получены для обеих дисциплин обслуживания требования после восстановления прибора.

В **третьей главе** найдены условия, при которых $\Psi(x)$ будет субэкспоненциальной функцией распределения. Сначала мы получаем результат для системы обслуживания с регенерирующим входящим потоком и надежным прибором, а затем переносим этот результат на систему с прерываниями.

В разделе 3.1 дается описание модели.

Рассматривается одноканальная система массового обслуживания с неограниченной очередью и регенерирующим входящим потоком $X(t)$, определенным на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ с фильтрацией $\{\mathscr{F}_t, t \geqslant 0\}$, к которой он адаптирован. Процесс $X(t)$ имеет непрерывные справа неубывающие кусочно-постоянные траектории ($X(0) = 0$).

Марковские моменты θ_j относительно фильтрации $\{\mathscr{F}_t, t \geqslant 0\}$ и случайные величины $\tau_j = \theta_j - \theta_{j-1}\,(j = 1, 2, \dots)$ называются соответственно моментами и периодами регенерации. Здесь мы считаем, что $X(t)$ — суммарное время обслуживания поступивших требований, и обозначаем $\gamma_n = X(\theta_n) - X(\theta_{n-1})$ и ξ_n — число скачков $X(t)$ на n-ом периоде регенерации. Предположим, что $\mathsf{E}\gamma_2 < \infty$ и $\mathsf{E}\tau_2 < \infty$, и обозначим $G(y) = \mathsf{P}(\gamma_2 \leqslant y)$, $g(x) = \int_x^\infty (1 - G(y))dy$, $G^I(x) = \frac{1}{\mathsf{E}\gamma_2} \int\limits_0^x (1 - G(y))dy$.

Процесс $W(t)$ — виртуальное время ожидания. Определим два вложенных процесса $W_n = W(\theta_n - 0)$ и $w_n = W(t_n - 0)$, где t_n — момент n-го скачка процесса $X(t)$. Для сходимости последовательности функций $\lim_{n\to\infty} \mathsf{P}(w_n \leqslant x) = F(x)$ при $n \to \infty$ достаточно потребовать апериодичность распределения ξ_2.

В работе [28](Афанасьева, Баштова, 2014) даны достаточные условия, при которых существуют пределы

$$\Phi(x) = \lim_{n\to\infty} \mathsf{P}(W_n \leqslant x),$$
$$\Psi(x) = \lim_{t\to\infty} \mathsf{P}(W(t) \leqslant x).$$

Более того, $\Psi(x)$, $\Phi(x)$ и $F(x)$ являются функциями распределения тогда и только тогда, когда коэффициент загрузки $\rho = \frac{\mathsf{E}\gamma_2}{\mathsf{E}\tau_2} < 1$.

Далее везде предполагаем, что предельные распределения существуют и $\rho < 1$.

Основная задача главы — найти асимптотическое поведение функций $\Psi(x)$, $\Phi(x)$ и $F(x)$ при $x \to \infty$. Будем считать, что случайный вектор (τ, γ, ξ) имеет такое же распределение, как и вектор $(\tau_2, \gamma_2, \xi_2)$.

В разделе 3.2 мы приводим основной результат. Обозначим класс субэкспоненциальных распределений $\mathscr{S}$.

Теорема 3. *Пусть $G^I(x) \in \mathscr{S}$ и при $y \to \infty$*

$$\mathsf{P}(\gamma > y + \tau) \sim \bar{G}(y). \tag{8}$$

Тогда при $x \to \infty$

$$\bar{\Phi}(x) \sim a^{-1} g(x), \tag{9}$$

где $a = \mathsf{E}\tau - \mathsf{E}\gamma$. В доказательстве используются построение вспомогательных мажорирующих систем и результаты [13](Боровков, 1972), [71](Veraverbeke, 1977) для случайных блужданий с приращениями, которые имеют тяжелые хвосты.

Чтобы получить аналогичную асимптотику для функции $\bar{\Psi}(x)$, мы воспользуемся результатами работы [32](Asmussen, 1999). Если в условиях теоремы 3 дополнительно потребовать, чтобы функция распределения $\mathsf{P}(\gamma > y + \tau)$ была локально постоянной (см. [13], обозначение $\mathscr{L}$), то асимптотика (9) верна для $\bar{\Psi}(x)$.

Введем функцию

$$v(x) = \frac{1}{\mathsf{E}\xi} \sum_{j=0}^{\infty} j\, \mathsf{P}(\gamma > x, \xi = j).$$

Теорема 4. *Пусть выполнены предположения теоремы 3, условие 6 и*

$$\lim_{t\to\infty} v(x)/g(x) = 0.$$

Тогда для $\bar{F}(x)$ имеет место (9).

Доказательство опирается на неравенства

$$\mathsf{P}(W_{r(n)-1} - \tau_{r(n)} > x) \leqslant \mathsf{P}(w_n > x) \leqslant \mathsf{P}(W_{r(n)-1} + \gamma_{r(n)} > x), \tag{10}$$

где $r(n)$ — номер периода регенерации, на котором пришло n-е требование. Далее, мы показываем, что верхняя и нижняя оценки асимптотически эквиваленты $\bar{\Phi}(x)$ при $x \to \infty$.

Приводится следствие, упрощающее проверку условий в случае, когда количество поступившей работы за период регенерации не зависит от длины этого периода.

Следствие 10. *Если $G(x) \in \mathscr{L}$, $G^I(x) \in \mathscr{S}$ и γ не зависит от τ, то* (9) *имеет место для функций $\bar{\Phi}(x)$ и $\bar{\Psi}(x)$.*

В разделе 3.3 мы рассматриваем стандартную систему $Reg/G/1/\infty$. Здесь мы предположим, что процесс $A(t)$, задающий число требований, поступивших в систему за $[0,t)$, является регенерирующим в смысле определения 3. Времена обслуживания $\{\eta_n\}_{n=1}^{\infty}$ образуют последовательность независимых одинаково распределенных случайных величин, не зависящих от $A(t)$. Тогда $X(t) = \sum\limits_{j=1}^{A(t)} \eta_j$ также регенерирующий процесс с теми же моментами регенерации, что и $A(t)$. Интенсивность потока $A(t)$ определяется соотношением $\lambda = \lim_{t\to\infty} \frac{A(t)}{t} = \frac{\mathsf{E}\xi}{\mathsf{E}\tau}$ (п.н.), а коэффициент загрузки $\rho = \lambda b < 1$, где $b = \mathsf{E}\eta_j$. Пусть $B(x)$ — функция распределения времени обслуживания. Введем вспомогательную случайную величину $\tilde{\xi}$, которая имеет распределение $\mathsf{P}(\tilde{\xi} = k) = \frac{k}{\mathsf{E}\xi}\mathsf{P}(\xi = k)$, $(k = 1, 2, \dots)$. Класс $\mathscr{S}^*$ — сильносубэкспоненциальные распределения.

Теорема 5. *Пусть $B(x) \in \mathscr{S}^*$, распределение ξ апериодично. Тогда верны следующие утверждения:*

(i) Если существует $c > b$ такое, что $\mathsf{P}(c\xi > x) = o(\bar{B}(x))$, то при $x \to \infty$

$$\bar{\Psi}(x) \sim \bar{\Phi}(x) \sim \frac{\lambda}{1 - \lambda b} \int_x^{\infty} \bar{B}(y)dy. \tag{11}$$

(ii) Если существует $c > b$ такое, что $\mathsf{P}(c\tilde{\xi} > x) = o(\bar{B}(x))$, то асимптотика (11) *справедлива и для функции $\bar{F}(x)$.*

Раздел 3.4 посвящен приложениям теоремы 5.

Так в примере 6 мы рассматриваем случай, когда в системе $Reg/G/1/\infty$ прибор ненадежен. Прибор может выходить из строя только во время обслуживания. Время рабочего состояния распределено экспоненциально с параметром ν, а время восстановления прибора имеет функцию распределения

$D(x)$, с ПЛС $d(s)$ и средним $d < \infty$. Если обслуживание требования было прервано, то после устранения прерывания требование обслуживается заново, причем новое время обслуживания не зависит от исходного (дисциплина $\mathsf{D2}$).

Следствие 11. *Пусть для системы $Reg/G/1$ с ненадежным прибором и дисциплиной $\mathsf{D2}$ функция распределения времени восстановления $D(x) \in \mathscr{S}^*$. Если $\mathsf{P}(r\xi > x) = o(\bar{D}(x))$ для некоторого $r > b_c$, то при $x \to \infty$*

$$\bar{\Psi}(x) \sim \bar{\Phi}(x) \sim \frac{\lambda(1 - b(\nu))}{(1 - \lambda b_c) b(\nu)} \int_x^\infty \bar{D}(y) dy.$$

Если $\mathsf{P}(r\tilde{\xi} > x) = o(\bar{D}(x))$, то эта асимптотика верна для $\bar{F}(x)$.

В примерах 7 и 8 рассмотрены случаи, когда в систему поступают ДСПП и полумарковски-модулированный поток. Показано, какой вид примет теорема 5 в этих случаях.

Благодарность. Автор выражает глубокую признательность своему научному руководителю доктору физико-математических наук, профессору Афанасьевой Ларисе Григорьевне за постановку задач и постоянное внимание к работе. Автор высоко ценит содействие и внимание, оказанное работе сотрудниками кафедры теории вероятностей механико-математического факультета МГУ имени М. В. Ломоносова.

Глава 1

Эргодическая теорема для системы обслуживания с ненадежным прибором

Данная глава посвящена определению необходимых и достаточных условий стабильности одноканальной системы обслуживания с регенерирующим входящим потоком и ненадежным прибором. Перед основным результатом мы дадим определение регенерирующего потока и приведем его основные свойства.

1.1 Определение и свойства регенерирующего случайного потока

Для описания системы обслуживания прежде всего задаётся входной поток требований.В настоящей работе мы опираемся на следующее определение входного потока.

Определение 1. Случайный процесс $\{X(t),\, t \geqslant 0\}$ называется *случайным потоком*, если все его траектории являются неубывающими и непрерывными справа функциями.

Существует и другое определение случайного потока, заданное через неубывающую последовательность моментов наступлений событий, однако эти определения эквивалентны [26, § 6, § 12](Хинчин, 1963). В теории массового обслуживания входящий поток обычно задается требованиями, поступающими в систему. Например, если принять за $X(t)$ количество требований, поступивших в отрезке $[0,t]$, то траектории $X(t)$ будут неубывающими непрерывными справа кусочно-постоянными функциями, принимающими целые неотрицательные значения. Под потоком мы будем понимать также и объем работы, которую мог бы выполнить прибор к моменту t, если бы в системе всегда были требования. Теории таких случайных потоков посвящена книга Хинчина [26].

В диссертации будут использоваться результаты, касающиеся регенерирующих процессов и потоков. В связи с этим представляется полезным привести основные определения и теоремы, относящиеся к ним. Нестрого говоря свойство регенерации означает независимость будущего поведения случайного процесса от прошлого в некоторые случайные моменты времени. Эволюция таких процессов представляет собой последовательность независимых циклов и изучается методами теории восстановления. В специальной литературе существует несколько путей формализации данного понятия [69](Thorisson, 1983), [30](Asmussen, 1987). В настоящей диссертации используется определение, предложенное в [22, Часть II, Глава 2](Кокс, Смит, 1967).

Пусть $(\Omega, \mathscr{F}, \mathsf{P})$ — основное вероятностное пространство с фильтрацией $\{\mathscr{F}_t, t \in T\}(T = [0, \infty)$ или $T = \mathbb{Z}^+)$, на котором определен случайный процесс $\{U(t), t \in T\}$, адаптированный к фильтрации $\mathscr{F}_t$. Процесс $U(t)$ принимает значения в измеримом фазовом пространстве $(\mathcal{X}, \mathscr{F}(\mathcal{X}))$.

Определение 2. Процесс $U(t)$ — называется *регенерирующим*, если существует неубывающая последовательность $\{\theta_j\}_{j=1}^{\infty}$ $(\theta_0 = 0)$ марковских моментов относительно фильтрации $\{\mathscr{F}_t, t \in T\}$ такая, что последовательность

$$\{\varkappa_i\}_{i=1}^{\infty} = \{\theta_i - \theta_{i-1}, U(t + \theta_{i-1}), t \in (0, \theta_i - \theta_{i-1}]\}_{i=1}^{\infty}$$

состоит из независимых одинаково распределенных случайных элементов на $(\Omega, \mathscr{F}, \mathsf{P})$ при $i > 1$.

Последовательность марковских моментов $\{\theta_i\}_{i=0}^{\infty}$ называется точками регенерации, а $\{\tau_i = \theta_i - \theta_{i-1}\}_{i=1}^{\infty}$ периодами регенерации. Определив регенерирующий случайный процесс $U(t)$, перейдем к введению регенерирующего случайного потока $X(t)$.

Определение 3. Случайный *поток* $\{X(t), t \in T\}$ назовем *регенерирующим*, если существует неубывающая последовательность $\{\theta_j\}_{j=1}^{\infty}$ $(\theta_0 = 0)$ марковских моментов относительно фильтрации $\{\mathscr{F}_t, t \in T\}$ такая, что последовательность

$$\{\varkappa_j\}_{j=1}^{\infty} = \{\theta_j - \theta_{j-1}, X(\theta_{j-1} + t) - X(\theta_{j-1}), t \in (0, \theta_j - \theta_{j-1}]\}_{j=1}^{\infty}$$

состоит из независимых случайных элементов, одинаково распределенных при $j > 1$.

В теории массового обслуживания входящий поток — это либо суммарная работа, либо число требований, поступивших к моменту t. В диссертации суммарная работа — $X(t)$, а число требований — $A(t)$. Обозначим $\gamma_i = X(\theta_i) - X(\theta_{i-1})$ и ξ_i — соответственно приращение и количество скачков процесса $X(t)$ за i-ый период регенерации. Если входящий поток $A(t)$

ординарный, то $\xi_i = A(\theta_i) - A(\theta_{i-1})$ — это число пришедших требований, а величина каждого скачка $X(t)$ — это время обслуживания требования. Для удобства обозначений введем случайный вектор (γ, τ, ξ), который совпадает по распределению с вектором $(\gamma_2, \tau_2, \xi_2)$.

Приведем некоторые результаты, которые будут использованы в дальнейшем. Будем предполагать, что начальная регенерация обязательно произойдет, то есть $\mathsf{P}(\tau_1 < \infty) = 1$.

Свойство 1. Пусть $X(t)$ — регенерирующий поток и $a := \mathsf{E}\gamma < \infty$, $\mu := \mathsf{E}\tau < \infty$. Тогда с вероятностью 1 существует

$$\lambda := \lim_{t\to\infty} \frac{X(t)}{t} = \frac{a}{\mu}.$$

Величина $\lambda < \infty$ называется *интенсивностью* потока $X(t)$.

Доказательство. Заметим, что $\{\theta_i, i \geqslant 1\}$ является общим процессом восстановления, $n(t)$ — количество восстановлений этого процесса на отрезке $[0, t]$. Так как $\mu < +\infty$, то по усиленному закону больших чисел для процессов восстановления ([10, Глава 1,§6] (Афанасьева, Булинская, 1980))

$$\lim_{t\to+\infty} \frac{n(t)}{t} = \mu \quad \text{п.н.}$$

Заметим, что

$$\sum_{i=1}^{n(t)-1} \gamma_i \leqslant X(t) \leqslant \sum_{i=1}^{n(t)} \gamma_i.$$

Так как по предположению $a < +\infty$ и $\mathsf{P}(\gamma_1 < +\infty) = 1$, получаем по усиленному закону больших чисел

$$\mathsf{E}\gamma = \lim_{t\to+\infty} \frac{\sum_{i=1}^{n(t)-1} \gamma_i}{n(t)} \leqslant \lim_{t\to+\infty} \frac{X(t)}{n(t)} \leqslant \lim_{t\to+\infty} \frac{\sum_{i=1}^{n(t)} \gamma_i}{n(t)} = \mathsf{E}\gamma \text{ п.н.}$$

Следовательно,

$$\lim_{t\to+\infty} \frac{X(t)}{t} = \lim_{t\to\infty} \frac{X(t)}{n(t)} \frac{n(t)}{t} = \frac{\mathsf{E}\gamma}{\mathsf{E}\tau} \quad \text{п.н.}$$

■

Определение 4. Случайная величина ξ имеет *апериодичное* распределение, если наибольший общий делитель $\{j : \mathsf{P}(\xi = j) > 0\} = 1$.

Свойство 2. Обозначим t_n момент поступления n-го требования ($n = 1, 2, \ldots$, $t_0 = 0$). Тогда $\{\zeta_n = t_n - t_{n-1}\}_{n=1}^{\infty}$ — последовательность интервалов между поступлениями требований. Пусть $d =$ Н.О.Д.$\{j : \mathsf{P}(\xi_n = j) > 0\}$. Если d = 1, то имеет место слабая сходимость $\{\zeta_{n+k}\}_{k=1}^{\infty}$ к стационарной метрически транзитивной последовательности $\{\hat{\zeta}_k\}_{k=1}^{\infty}$ при $n \to \infty$.

Доказательство этого свойства опирается на основную теорему восстановления для непериодических решетчатых случайных процессов восстановления и приведено в [28] (Афанасьева, Баштова, 2014).

Далее условие $d = 1$ предполагается выполненным.

Приведем ещё одно определение регенерирующего потока, которое наглядно демонстрирует связь определений 2 и 3.

Определение 5. Случайный *поток* $\{X(t), t \in T\}$ — *регенерирующий*, если существует неубывающая последовательность $\{\theta_j\}_{j=1}^{\infty} (\theta_0 = 0)$ марковских моментов относительно фильтрации $\{\mathscr{F}_t, t \in T\}$ такая, что процесс

$$\hat{X}(t) = X(t) - X(\theta_{n(t)})$$

является регенерирующим, здесь $n(t) = \max\{k : \theta_k < t\}$.

Условие 1. Пусть для случайной величины τ существуют $n \geqslant 1$ и неотрицательная функция $f(x)$ такие, что $\int\limits_{\mathbb{R}} f(x)dx > 0$ и

$$\mathsf{P}(\hat{\tau}_1 + \dots + \hat{\tau}_n \in A) \geqslant \int_A f(x)dx, \quad A \in \mathcal{B},$$

где $\hat{\tau}_1, \dots, \hat{\tau}_n$ независимые случайные величины, распределенные как τ. В зарубежной литературе такие распределения называются — *spread out distributions* (см. [30, Глава 7](Asmussen, 1987)).

Определение 6. Последовательность случайных величин $\{\tau_n,\, n \geqslant 1\}$, определенная на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$, сходится *по вариации* к случайной величине τ ($\tau_n \xrightarrow[n\to\infty]{tv} \tau$), если при $n \to \infty$

$$\sup_{A \in \mathscr{F}} |\,\mathsf{P}(\tau_n \in A) - \mathsf{P}(\tau \in A)| \to 0$$

Свойство 3. Пусть $U(t)(X(t))$ регенерирующий процесс (регенерирующий поток) и распределение периода регенерации τ удовлетворяет условию 1. Тогда

$$U(t+z) \xrightarrow[t\to\infty]{tv} \tilde{U}(z),$$

$$X(t+z) \xrightarrow[t\to\infty]{tv} \tilde{X}(z),$$

где $\tilde{U}(z)(\tilde{X}(z))$ стационарный процесс (поток со стационарными приращениями). Если период регенерации τ имеет решетчатое распределение, то достаточно потребовать чтобы распределение τ было апериодичным.

Свойство 4. Если длина первого периода регенерации имеет распределение

$$\text{в непрерывном случае} \qquad \mathsf{P}(\tau_1 > x) = (\mathsf{E}\tau)^{-1} \int_x^\infty \mathsf{P}(\tau > y)dy,$$
$$\text{в решетчатом случае} \qquad \mathsf{P}(\tau_1 = k) = (\mathsf{E}\tau)^{-1}\, \mathsf{P}(\tau > k),$$

то процесс $U(t)$ стационарный, а поток $X(t)$ является процессом со стационарными приращениями.

Данные два свойства доказаны в книге [70, Глава 10](Thorisson, 2000).

Теорема (Основная теорема восстановления)**.** *Рассматривается общий непрерывный (т.е. нерешетчатый) процесс восстановления* $\{\theta_n, n \geqslant 0\}$*, образованный последовательностью независимых случайных величин* $\{\tau_n, n \geqslant 1\}$*, то есть* $\theta_n = \sum\limits_{k=0}^{n} \tau_k$*. Функция восстановления —* $H(t) = \sum\limits_{k=0}^{\infty} \mathsf{P}(\theta_k \leqslant t)$*. Пусть* $z(t)$ *— функция ограниченной вариации, интегрируемая на* $[0, \infty)$*. Тогда*

$$\lim_{t\to\infty} \int_0^t z(t-x)dH(x) = \begin{cases} \frac{1}{\mathsf{E}\tau_2} \int_0^\infty z(s)ds, & \textit{если}\, \mathsf{E}\tau_2 < \infty, \\ 0, & \textit{если}\, \mathsf{E}\tau_2 = \infty. \end{cases}$$

Теорема (Блекуэлл, непрерывный случай)**.** *Для функции восстановления* $H(t)$ *справедлива асимптотика*

$$\lim_{t\to\infty} (H(t+\alpha) - H(t)) = \frac{\alpha}{\mathsf{E}\tau_2}$$

при любом фиксированном $\alpha > 0$*. Если* $\mathsf{E}\tau = \infty$*, то правая часть равна нулю.*

Теорема (Блекуэлл, дискретный случай)**.** *Рассматривается общий дискретный процесс восстановления* $\{\theta_n, n \geqslant 0\}$ *с шагом* h*, образованный последовательностью собственных случайных величин* $\{\tau_n, n \geqslant 1\}$*. Если распределение* τ_n *является апериодическим, то для функции восстановления* $H(t)$ *справедлива асимптотика*

$$H(nh) - H((n-l)h) \to \frac{lh}{\mathsf{E}\tau_2}, \quad n \to +\infty.$$

Если $\mathsf{E}\tau = +\infty$*, то правая часть равна нулю.*

Доказательство этих теорем в случае $\mathsf{E}\tau < \infty$ содержится в [14](Боровков, 1999). В [10](Афанасьева, Булинская, 1980) приведена схема доказательства для случая $\mathsf{E}\tau = \infty$, а также показано, что две эти теоремы эквивалентны.

Для того чтобы установить существование предельного распределения у регенерирующего процесса, мы будем пользоваться теоремой из [67](Смит, 1955). Для регенерирующего процесса $U(t)$ введем функцию $\Phi_A(t)$ следующим образом

$$\Phi_A(t-\kappa_t) := \mathsf{P}(U(t)\in A|\mathscr{F}_{\leqslant\kappa_t}),$$

где $\kappa_t = \max\{\theta_j : \theta_j < t\}$ — последний момент регенерации, произошедший до t. Заметим, что κ_t не является марковским моментом. Найти формальное построение σ-алгебры $\mathscr{F}_{\leqslant\kappa_t}$ можно в [10, Раздел 2, §3](Афанасьева, Булинская, 1980).

Нам понадобится определение непосредственной интегрируемости по Риману. Для функции $f(x)(x\in\mathbb{R})$ введем верхние и нижние суммы $\overline{\sigma}$ и $\underline{\sigma}$. Пусть $\overline{m}_n$ и $\underline{m}_n$ — максимум и минимум функции $f(x)$ на отрезке $[(n-1)h, nh]$ для фиксированного $h>0$. Обозначим $\overline{\sigma} = \sum\limits_{n=-\infty}^{\infty} h\overline{m}_n$ и $\underline{\sigma} = \sum\limits_{n=-\infty}^{\infty} h\underline{m}_n$.

Определение 7. Функция $f(x)$ называется *непосредственно интегрируемой по Риману*, если

1. ряды $\overline{\sigma}$ и $\underline{\sigma}$ сходятся абсолютно,
2. $\overline{\sigma} - \underline{\sigma} \to 0$ при $h\to 0$.

Теорема (Смит, непрерывный случай)**.** *Пусть* $\{U(t), t\geqslant 0\}$ *— непрерывный регенерирующий процесс,* $\Phi_A(t)(t\in[0,\infty))$ *— борелевская функция,* $F(t) = \mathsf{P}(\tau_2\leqslant t)$. *Если выполнены условия*

1. $\mathsf{P}(\tau_1<\infty)=1$,
2. *функция* $\Psi_A(t) = \Phi_A(t)(1-F(t))$ *непосредственно интегрируема по Риману,*

то существует

$$\lim_{t\to\infty}\mathsf{P}(U(t)\in A) = \begin{cases} \frac{1}{\mathsf{E}\tau_2}\int_0^\infty \Psi_A(s)ds, & \textit{если}\, \mathsf{E}\tau_2<\infty,\\ 0, & \textit{если}\, \mathsf{E}\tau_2=\infty.\end{cases}$$

Доказательство этой теоремы для случая $\mathsf{E}\tau_2<\infty$ приведено, например, в [10](Афанасьева, Булинская, 1980). Результаты для случая $\mathsf{E}\tau_2=\infty$ получены в [69](Thorisson, 1983).

Как показано в [22](Кокс, Смит, 1967), функция $\Psi_A(t)$ удовлетворяет условию 2, если

A. функция $\Psi_A(t)$ — интегрируемая и не возрастающая на $(0,\infty)$.

B. функция $\Psi_A(t)$ имеет ограниченную вариацию на любом конечном интервале времени и $\mathsf{E}\tau<\infty$.

C. случайная величина τ_2 удовлетворяет условию 1 и $\mathsf{E}\tau_2 < \infty$.

Определение 8. Процесс $U(t)$ называется *стохастически ограниченным*, если

$$\lim_{x\to\infty} \limsup_{t\to\infty} \mathsf{P}(U(t) > x) = 0.$$

Определение 9. Процесс $U(t)$ называется *эргодическим*, если он имеет собственное предельное распределение, не зависящее от начального состояния системы. Другими словами, для любого $x \in \mathbb{R}$ существует предел

$$\lim_{t\to\infty} \mathsf{P}(U(t) \leqslant x) = F_U(x),$$

и $F_U(x)$ является функцией распределения, не зависящей от состояния системы в момент 0.

В западной литературе данной свойство называется стабильностью.

1.2 Примеры регенерирующих потоков

Класс регенерирующих потоков включает в себя широкий круг практически важных потоков. Ниже приведены некоторые примеры.

1.2.1 Дважды стохастический пуассоновский поток (ДСПП)

Приведём два определения ДСПП, эквивалентность которых доказана в [49] (Grandell, 1976).

Определение 10. Пусть на одном вероятностном пространстве задан случайный поток $\{\Lambda(t,\omega), t \geqslant 0\}$, $\Lambda(0,\omega) = 0$, и стандартный пуассоновский процесс $\{A(t,\omega), t \geqslant 0\}$ (то есть пуассоновский процесс с интенсивностью 1), не зависящий от $\{\Lambda(t,\omega), t \geqslant 0\}$. *Дважды стохастический пуассоновский поток* $\{X(t,\omega), t \geqslant 0\}$ определяется

$$X(t,\omega) := A(\Lambda(t,\omega),\omega)$$

и $\Lambda(t)$ называется *ведущим потоком* ДСПП.

Приведённое выше определение демонстрирует, что ДСПП получается из стандартного пуассоновского процесса при помощи случайной замены времени.

Определение 11. Пусть $\{X(t,\omega), t \geqslant 0\}$ — случайный процесс с неубывающими траекториями, принимающий значения в $\mathbb{Z}_+$. Если существует случайный поток $\{\Lambda(t,\omega), t \geqslant 0\}$, $\Lambda(0,\omega) = 0$ такой, что при условии $\Lambda(t,\omega) \equiv \Lambda_0(t)$ процесс $X(t,\omega)$ является пуассоновским с ведущей функцией $\Lambda_0(t)$, то $X(t,\omega)$ называется *дважды стохастическим пуассоновским потоком.*

Утверждение 1. *Если $\Lambda(t)$ — регенерирующий поток с точками регенерации $\{\theta_k, k \geqslant 0\}$, то $X(t)$ — регенерирующий поток с точками регенерации $\{\theta_k, k \geqslant 0\}$. Обозначим $\tau_k = \theta_k - \theta_{k-1}$.*

Доказательство. Докажем, что $X(t)$ удовлетворяет определению 3. Пусть σ-алгебра $\mathcal{G} := \sigma\{\Lambda(t), t \geqslant 0\}$. Для краткости рассмотрим одномерные распределения случайных элементов $x_k(t) = (X(t+\theta_{k-1}) - X(\theta_{k-1}))\mathbb{I}(t \in (0, \tau_k])$. Справедливо

$$\begin{aligned}\mathsf{P}(x_k(t) = i, x_l(u) = j) = \mathsf{E}\,\mathsf{P}(x_k(t) = i, x_l(u) = j \,|\, \mathcal{G}) = \\ = \mathsf{E}\left(\frac{(\Lambda(\theta_{k-1}+t) - \Lambda(\theta_{k-1}))^i}{i!} e^{-(\Lambda(\theta_{k-1}+t)-\Lambda(\theta_{k-1}))}\mathbb{I}_{[0,\tau_k]}(t)\times\right. \\ \left.\times\frac{(\Lambda(\theta_{l-1}+u) - \Lambda(\theta_{l-1}))^j}{j!} e^{-(\Lambda(\theta_{l-1}+u)-\Lambda(\theta_{l-1}))}\mathbb{I}_{[0,\tau_l]}(u)\right) = \\ = \mathsf{P}(x_k(t) = i)\,\mathsf{P}(x_l(u) = j).\end{aligned} \quad (1.1)$$

Выше был использован тот факт, что в силу определения 11 верно

$$\begin{aligned}\mathsf{P}(x_k(t) = i \,|\, \mathcal{G}) = \mathsf{P}\big((X(\theta_{k-1}(\omega)+t, \omega) - X(\theta_{k-1}(\omega), \omega))I_{[0,\tau_k(\omega)]}(t) = i \,|\, \mathcal{G}\big) = \\ = \frac{(\Lambda(\theta_{k-1}+t) - \Lambda(\theta_{k-1}))^i}{i!} e^{-(\Lambda(\theta_{k-1}+t)-\Lambda(\theta_{k-1}))} I_{[0,\tau_k]}(t).\end{aligned}$$

С учётом определения 3 и того, что $\Lambda(t)$ — регенерирующий поток, соответствующие сомножители в (1.1) независимы, что доказывает данное утверждение. □

Используя определение 11, нетрудно проверить, что ДСПП $X(t)$ обладает следующими свойствами.

1. $\mathsf{E}X(t) = \mathsf{E}\Lambda(t)$.
2. $\mathsf{D}\,X(t) = \mathsf{D}\,\Lambda(t) + \mathsf{E}\Lambda(t)$.
3. В случае, когда $\Lambda(t)$ — регенерирующий поток, интенсивность ДСПП совпадает с интенсивностью ведущего потока

$$\lambda = \lim_{t\to+\infty}\frac{\Lambda(t)}{t} = \frac{\mathsf{E}(\Lambda(\theta_n) - \Lambda(\theta_{n-1}))}{\mathsf{E}\tau_n}. \quad (1.2)$$

Если ведущий поток представляется в виде $\Lambda(t) = \int_0^t \lambda(u)\,du$, где $\{\lambda(t,\omega), t \geqslant 0\}$ — неотрицательный и ограниченный с вероятностью 1 случайный процесс, то $\lambda(t)$ называется *случайной интенсивностью* потока $X(t)$.

Утверждение 2. *Если $\lambda(t)$ — регенерирующий процесс с точками регенерации $\{\theta_k, k \geqslant 0\}$, то $\Lambda(t)$ — регенерирующий поток с точками регенерации $\{\theta_k, k \geqslant 0\}$.*

Доказательство. Рассмотрим последовательность случайных элементов

$$\varphi_k(\omega) = (\lambda(\theta_{k-1} + t, \omega), \tau_k(\omega), t \in [0, \tau_k)), \quad k \geqslant 1.$$

и

$$\chi_k(\omega) = (\Lambda(\theta_{k-1} + t, \omega) - \Lambda(\theta_{k-1}, \omega), \tau_k(\omega), t \in [0, \tau_k]), \quad k \geqslant 1.$$

Пусть отображение F, такое что $F(\varphi_k) = \chi_k$, $k \geqslant 1$. Заметим, что $\Lambda(\theta_{k-1} + t) - \Lambda(\theta_{k-1}) = \int_0^t \lambda(\theta_{k-1} + u)\, du$. Из определения 3 следует независимость и одинаковая распределённость φ_k, $k \geqslant 1$. В силу измеримости F случайные элементы χ_k, $k \geqslant 1$, будут также независимы и одинаково распределены. Аналогично проверяется независимость от χ_0. По определению 3 поток $\Lambda(t)$ является регенерирующим с моментами регенерации $\{\theta_k\}$. □

1.2.2 Поток с интенсивностью случайной амплитуды

Случайные амплитуды возникают, когда известен общий закон изменения интенсивности входного потока в течение некоторого периода, но на амплитуду изменения оказывают влияние дополнительные случайные факторы.

Строгое определение выглядит следующим образом. Пусть $v(t)$ — периодическая интегрируемая неотрицательная функция с периодом T, $\{\eta_n, n \geqslant 0\}$ — последовательность независимых одинаково распределенных случайных величин ($\eta_n \geqslant 0$ п.н.). Случайную интенсивность ДСПП определим равенством

$$\lambda(nT + t, \omega) := v(t)\eta_n(\omega), \quad n \geqslant 0.$$

При таком определении величины η_n можно интерпретировать как случайные амплитуды, которые постоянны на n-м периоде. Моменты регенерации неслучайны и задаются последовательностью $\{nT, n \geqslant 0\}$, приращение ДСПП за n-й период регенерации равно $\eta_n \int_0^T v(t)dt$, интенсивность потока

$$\lambda = \frac{\int_0^T v(t)\, dt \mathsf{E}\eta_1}{T}. \tag{1.3}$$

1.2.3 Поток со случайными периодами

Пусть, как и раньше, $v(t)$ — периодическая интегрируемая неотрицательная функция с периодом T, $\{\eta_n, n \geqslant 1\}$ — последовательность независимых одинаково распределенных случайных величин ($\eta_n \geqslant 0$ п.н.). Случайную интенсивность определим следующим образом

$$\lambda(t, \omega) := v(\gamma_t(\omega)),$$

где γ_t — недоскок процесса восстановления, образованного последовательностью $\{\eta_n, n \geqslant 1\}$, до момента t. Моментами регенерации ДСПП являются моменты восстановлений процесса $\{S_n, n \geqslant 0\}$, $S_n = \sum_{k=1}^{n} \eta_k$, приращение за n-й период регенерации равно $\int_{\eta_{n-1}}^{\eta_n} v(\gamma_t)\,dt$, интенсивность потока

$$\lambda = \frac{\mathsf{E}\int_0^{\eta_1} v(t)\,dt}{\mathsf{E}\eta_1}. \tag{1.4}$$

Приведём теперь примеры дважды стохастических пуассоновских процессов, возникающих в классических системах массового обслуживания.

1.2.4 Поток потерянных требований

Рассмотрим r-канальную систему, в которую поступает пуассоновский поток требований интенсивности λ. Время обслуживания каждого требования имеет показательное распределение с параметром ν. Заявка, поступающая в систему, в которой уже находятся j требований, присоединяется к очереди с вероятностью f_j и теряется с вероятностью $1 - f_j$, $f_j \in [0, 1]$. Считаем, что $f_j = 1$, $0 \leqslant j < r$, то есть требование поступает на обслуживание с вероятностью 1, если есть свободный прибор. Пусть $Q(t)$ — число требований в системе в момент t, а $L(t)$ - число потерянных требований за $[0, t]$. Покажем, что $L(t)$ является ДСПП со случайной интенсивностью $\lambda(1 - f_{Q(t)})$.

Проверим определение 11. Зафиксируем траекторию $Q(t, \omega_0) = Q_0(t)$. Возьмём n непересекающихся полуинтервалов $(s_i, t_i]$, $1 \leqslant i \leqslant n$. Без ограничения общности можно считать, что $(s_i, t_i]$ — участки постоянства зафиксированной траектории $Q_0(t)$. Пусть длина интервала $(s_i, t_i]$ равна Δ_i, а $Q_0(t)$ принимает на нём значение k_i. Введём события

$$A_i := \{\text{На } (s_i, t_i] \text{ пришло } m_i \text{ требований, и все они были потеряны}\}$$

и

$$B_i := \{\text{Все требования, пришедшие на } (s_i, t_i], \text{ были потеряны}\}.$$

Тогда

$$\begin{aligned}
P(L(t_1) - L(s_1) = m_1, \ldots, L(t_n) - L(s_n) = m_n \,|\, Q(t) = Q_0(t), t \geqslant 0) = \\
= P\left(\bigcap_{i=1}^{n} A_i \,\middle|\, \bigcap_{i=1}^{n} B_i, Q(t) = Q_0(t), t \geqslant 0\right) = \\
= \frac{\prod\limits_{i=1}^{n} \frac{(\lambda\Delta_i)^{m_i}}{m_i!} e^{-\lambda\Delta_i}(1 - f_{k_i})^{m_i}}{\prod\limits_{i=1}^{n} \sum\limits_{l=0}^{+\infty} \frac{(\lambda\Delta_i)^{l}}{l!} e^{-\lambda\Delta_i}(1 - f_{k_i})^{l}} = \frac{\prod\limits_{i=1}^{n} \frac{(\lambda\Delta_i)^{m_i}}{m_i!} e^{-\lambda\Delta_i}(1 - f_{k_i})^{m_i}}{\prod\limits_{i=1}^{n} e^{-\lambda\Delta_i f_{k_i}}} = \\
= \prod_{i=1}^{n} \frac{(\lambda\Delta_i(1 - f_{k_i}))^{m_i}}{m_i!} e^{-\lambda\Delta_i(1 - f_{k_i})}.
\end{aligned}$$

Определение 11 проверено.

Моменты регенерации потока потерянных требований

$$\theta_k = \inf\{t_n, n \geqslant 1 \colon t_n > \theta_{k-1}, Q(t_n) = 0\}, \quad \theta_0 = 0,$$

где t_n — момент n-го скачка входного потока.

С другой стороны, в работе Л. Г. Афанасьевой [2](Афанасьева, 1966) показано, что в такой системе поток потерянных требований является полумарковским процессом. Таким образом, ДСПП в некоторых случаях является полумарковским процессом.

Если же $f_k = 0, k \geqslant r$, то согласно теореме Хинчина [26](Хинчин, 1963) поток потерянных требований образует процесс восстановления. Следовательно, в этом случае ДСПП будет являться процессом восстановления.

1.2.5 Марковски-модулированный поток (ММП)

Эти потоки рассматривались в [31](Asmussen, 1991) и они образуют важный подкласс ДСПП. Пусть $\{Y(t), t \geqslant 0\}$ — эргодическая цепь Маркова с конечным или счётным множеством состояний, а $\{\lambda_k, \lambda_k < C, k \geqslant 1\}$ — совокупность неотрицательных чисел.

Определение 12. ДСПП называется *марковски-модулированным потоком* (Markov-modulated flow) с управляющим процессом $Y(t)$, если его случайная интенсивность имеет вид

$$\lambda(t) = \sum_{k=1}^{+\infty} \lambda_k \mathbb{I}\{Y(t) = k\}.$$

Цепь Маркова $Y(t)$ является регенерирующим процессом, а точки регенерации определяются моментами достижения некоторого фиксированного состояния, скажем состояния 1. Другими словами, если $\{t_n, n \geqslant 1\}$ — последовательность моментов изменения состояний цепи Маркова $Y(t)$, то

$$\theta_k = \inf\{t_n, n \geqslant 1 \colon t_n > \theta_{k-1}, Y(t_n) = 1\}, \quad \theta_0 = 0. \tag{1.5}$$

Из утверждений 1 и 2 следует, что $X(t)$ также будет регенерирующим с моментами регенерации $\{\theta_k\}$. Интенсивность потока

$$\lambda = \sum_{k=1}^{+\infty} \lambda_k \pi_k, \tag{1.6}$$

где $\{\pi_k\}$ — стационарное распределение цепи $Y(t)$.

В силу сделанных предположений $\lambda < +\infty$. Среднее значение периода регенерации μ находится из соотношения $\pi_1 = (\alpha_1 \mathsf{E}\tau_k)^{-1}$, то есть

$$\mu = \mathsf{E}\tau_k = (\alpha_1 \pi_1)^{-1}. \tag{1.7}$$

Здесь α_1 — параметр экспоненциального распределения, задающего время пребывания управляющей цепи Маркова $Y(t)$ в состоянии 1. Заметим также, что если $\mathsf{P}(Y(0) = 1) = 1$, то $\{\theta_k\}$ образует простой процесс восстановления, а если $\mathsf{P}(Y(0) = j) = \pi_j$, $j \geqslant 0$, то этот процесс стационарный. Среднее число требований, поступивших за период регенерации, определяется соотношением

$$\mathsf{E}\xi_k = \lambda \mathsf{E}\tau_k = \frac{\lambda}{\alpha_1 \pi_1}. \tag{1.8}$$

Заметим, что потоки из разделов 1.2.2 и 1.2.3 не являются ММП.

1.2.6 Поток Льюиса

Потоки этого рода часто возникают в приложениях. Предположим, что в бесконечноканальную систему S поступает регенерирующий поток с моментами регенерации $\{\theta_n\}_{n=1}^{\infty}$. Времена обслуживания — независимые одинаково распределённые случайные величины. В течение времени пребывания в системе каждое требование независимо от других порождает пуассоновский поток интенсивности ν. Суммарный поток, генерируемый требованиями, находящимися на обслуживании в S, называется *потоком Льюиса*. Если $Q(t)$ — процесс числа требований в системе S, то поток Льюиса является ДСПП со случайной интенсивностью

$$\lambda(t) = \nu Q(t).$$

Точки регенерации потока

$$T_k = \inf\{\theta_n, n \geqslant 1 \colon \theta_n > T_{k-1}, Q(\theta_n) = 0\}, \quad T_0 = 0.$$

Интенсивность

$$\lambda = \nu \mathsf{E} Q, \tag{1.9}$$

где Q — случайная величина, которая имеет распределение процесса $Q(t)$ в стационарном режиме.

Поток Льюиса, вообще говоря, не является ММП. Он будет таковым, если входной поток — пуассоновский, а времена обслуживания имеют показательное распределение.

В качестве примера возможного приложения рассмотрим модель работы авиадиспетчера. Пусть $Q(t)$ — число воздушных судов в зоне наблюдения. Времена обслуживания интерпретируются, как времена нахождения судна в этой зоне. Пребывая в зоне наблюдения, каждое воздушное судно посылает запросы авиадиспетчеру на обслуживание. Эти запросы образуют входной поток, который и будет потоком Льюиса.

1.2.7 Марковский поток поступлений (МПП)

Рассмотрим эргодическую цепь Маркова $\{Y(t), t \geqslant 0\}$, принимающую значения в $\mathbb{Z}_+$ с непрерывными справа траекториями. Обозначим $\{t_j, j \geqslant 0\}$, $t_0 = 0$, неубывающую последовательность моментов скачков процесса $Y(t)$. Случайные величины $Y_j := Y(t_j)$, $j \geqslant 0$, образуют вложенную цепь Маркова с дискретным временем. Пусть $N(t) := \max\{j \geqslant 0 \colon t_j \leqslant t\}$. На исходном вероятностном пространстве зададим семейство $\varkappa := \{\{\varkappa_{ij}^{(n)}\}, n \in \mathbb{Z}; i, j \geqslant 0\}$ независимых последовательностей (при фиксированных i, j последовательность по n), состоящих из независимых одинаково распределённых случайных величин, принимающих значения в $\mathbb{Z}_+$. Математические ожидания равномерно ограничены $\mathsf{E}\varkappa_{ij}^{(n)} < C < +\infty$. Предполагается также, что семейство $\varkappa$ не зависит от $Y(t)$.

Определение 13. Случайный поток

$$X(t) = \sum_{j=1}^{N(t)} \varkappa_{Y_{j-1}Y_j}^{(j)}$$

называется *марковским потоком поступлений* (Markov arrival process) с управляющим процессом $Y(t)$.

МПП генерирует заявки в моменты перехода управляющей цепи. Поток $X(t)$ не является, вообще говоря, ДСПП, однако он регенерирующий с моментами регенерации $\{\theta_k\}$, определяемыми (1.5). Интенсивность потока задается соотношением

$$\lambda = \sum_{i=0}^{+\infty} \sum_{\substack{j=0 \\ j \neq i}}^{+\infty} \frac{\pi_i}{m_i} P_{ij} \mathsf{E}\varkappa_{ij}, \tag{1.10}$$

где $\{P_{ij}\}$ — матрица переходных вероятностей вложенной цепи $\{Y_j\}$, $\pi_i = \lim\limits_{t\to\infty} \mathsf{P}(Y(t) = i)$, m_i среднее время нахождения $Y(t)$ в состоянии i. Среднее значение периода регенерации определяется формулой (1.7), а для $\mathsf{E}\xi_k$ справедливо соотношение (1.8).

ММП образуют подкласс МПП. Для установления соответствия рассмотрим вспомогательную цепь Маркова $\widehat{Y}(t)$ с состояниями $\{(i, s), i, s \geqslant 0\}$, где i обозначает состояние цепи $Y(t)$, а s — количество скачков пуассоновского процесса за время пребывания ММП в состоянии i до момента t включительно. Возможными переходами $\widehat{Y}(t)$ являются $(i,0) \to (i,1) \to (i,2) \to \dots$ Более того, из любого состояния (i,s) возможен переход в $(j,0)$.

Пусть λ_i — интенсивность пуассоновского потока во время пребывания $Y(t)$ в состоянии i, $Q = \{q_{ij}\}$ — инфинитезимальная матрица $Y(t)$, тогда инфинитезимальные параметры $\widehat{Y}(t)$ имеют вид

$$q_{(i,0),(i,0)} = -\lambda_i + q_{ii}, \quad q_{(i,s),(i,s+1)} = \lambda_i, \quad q_{(i,s),(j,0)} = q_{ij}, \quad \text{где} \quad q_{ii} = -\sum_{i\neq j} q_{ij}.$$

Скачки МПП возникают в моменты перехода $\widehat{Y}(t)$ из состояния (i,s) в состояние $(i,s+1)$.

1.2.8 Полумарковский поток (ПМП)

Пусть $\{R(t), t \geqslant 0\}$ полумарковский процесс, принимающий значения в $\mathbb{Z}_+$. Вложенная цепь Маркова $\{\zeta_n, n \geqslant 1\}$ эргодична и имеет матрицу перехода $P = \{P_{ij}\}$. Обозначим $F_{ij}(x)$ функцию распределения времени пребывания процесса в состоянии i при условии, что переход будет осуществлён в состояние j. Пусть $m_{ij} := \int_0^{+\infty} x\,dF_{ij}(x)$, $F_i(x) := \sum_{j=0}^{+\infty} F_{ij}(x)$, $m_i := \int_0^{+\infty} x\,dF_i(x)$, и предположим, что $0 < m_{ij} < m < +\infty$, для всех i, j.

Дополнительно предположим, что семейство случайных потоков со стационарными приращениями $Z = \{\{Z_{ij}^{(n)}(t), t \geqslant 0\}, n \in \mathbb{Z}_+; i, j \geqslant 0\}$ не зависит от $R(t)$ и задано на исходном вероятностном пространстве. При фиксированных i, j последовательность $\{\{Z_{ij}^{(n)}(t), t \geqslant 0\}, n \in \mathbb{Z}_+\}$ состоит из независимых одинаково распределённых случайных элементов. Потоки $Z_{ij}^{(n)}(t)$ принимают значения в $\mathbb{Z}_+$. Предположим, что интенсивности потоков ограничены

$$\lambda_{ij} = \lim_{t\to+\infty} \frac{Z_{ij}^{(n)}(t)}{t} < \lambda < +\infty.$$

Обозначим $\{t_j, j \geqslant 0\}$ неубывающую последовательность моментов перехода процесса $R(t)$, $t_0 = 0$, так что $\zeta_j = R(t_j)$, $\zeta_0 = R(0)$. Введём процесс $N(t) = \max\{n \geqslant 0\colon t_n < t\}$.

Обозначим Δ_k приращение процесса $Z_{\zeta_{k-1}\zeta_k}^{(k)}(t)$ на интервале $[t_{k-1}, t_k)$, $k \geqslant 1$. Введём также приращение $\Delta(t)$ процесса $Z_{\zeta_{N(t)}\zeta_{N(t)+1}}^{(N(t)+1)}(t)$ на интервале $[t_{N(t)}, t)$. Определим процесс

$$X_1(t) := \sum_{k=1}^{N(t)} \Delta_k + \Delta(t). \tag{1.11}$$

Процесс определён для $t \geqslant 0$, имеет неубывающие траектории, $X_1(0) = 0$, $X_1(t) \geqslant 0$. Во время пребывания управляющего процесса $R(t)$ в состоянии

i, при условии перехода в состояние j, скачки процесса $X_1(t)$ совпадают со скачками соответствующего потока $Z_{ij}^{(n)}(t)$.

Определение 14. Если $Z_{ij}^{(n)}(t)$ является пуассоновским процессом с интенсивностью λ_{ij}, то $X_1(t)$ называется *полумарковски-модулированным потоком* (ПММП).

В данном случае поток $X_1(t)$ является аналогом ММП [30](S. Asmussen, 1987). ПММП является ДСПП со случайной интенсивностью

$$\lambda(t) = \sum_{i,j=0}^{+\infty} \lambda_{ij} \mathbb{I} A_{ij}^t,$$

где $A_{ij}^t = \bigcup_{n=0}^{+\infty} \{t \in [t_n, t_{n+1}), \zeta_n = i, \zeta_{n+1} = j\}$.

Также можно предположить, что

$$\lambda(t) = \sum_{i=0}^{+\infty} \lambda_i \mathbb{I}\{R(t) = i\}.$$

Рассмотрим вспомогательный полумарковский процесс $\widehat{R}(t)$ с множеством возможных состояний $\{(i,j)\}$, вероятностями перехода $\widehat{P}_{(i,j),(k,m)} = \delta_{jk} P_{jm}$ и функцией распределения времени пребывания $F_{(i,j)}^{(j,m)}(x) = F_{jm}(x)$. Последовательности моментов скачков данного процесса и процесса $R(t)$ совпадают. Следовательно, $X_1(t)$ является ДСПП со случайной интенсивностью

$$\lambda(t) = \sum_{(i,j)} \lambda_{ij} \mathbb{I}\{\widehat{R}(t) = (i,j)\}.$$

Аналогичным образом строятся процессы $\widehat{Z}_{ij}^{(n)}(t)$. Если интенсивности λ_{ij} не зависят от j, тогда интенсивность потока $X_1(t)$

$$\lambda_{X_1} = \lim_{t \to +\infty} \frac{X_1(t)}{t} = \sum_{i=0}^{+\infty} \lambda_i \pi_i, \qquad (1.12)$$

где $\pi_i = \lim_{t\to+\infty} \mathsf{P}(R(t) = i)$. В общем случае вместо (1.12) имеем

$$\lambda_{X_1} = \sum_{(i,j)} \lambda_{ij} \widehat{\pi}_i, \qquad (1.13)$$

где $\pi_{ij} = \lim_{t\to+\infty} \mathsf{P}(\widehat{R}(t) = (i,j))$. Пусть $\{\widehat{\pi}_{ij}^0\}$ — стационарное распределение вложенной цепи Маркова процесса $\widehat{R}(t)$, тогда $\widehat{\pi}_{ij}^0 = \pi_i^0 P_{ij}$, где $\{\pi_i^0\}$ — стационарное распределение $\{\zeta_n\}$. Общеизвестно [32](S. Asmussen, 1999), что

$$\pi_i = \frac{\pi_i^0 m_i}{\sum_{j=0}^{+\infty} \pi_j^0 m_j}, \quad \pi_i^0 = \frac{\pi_i \sum_{j=0}^{+\infty} \pi_j^0 m_j}{m_i}.$$

Используя данные соотношения, получаем

$$\widehat{\pi}_{ij}^0 = \frac{\pi_i^0 P_{ij} m_{ij}}{\sum_{(k,l)} \pi_k^0 P_{kl} m_{kl}} = \frac{\pi_i P_{ij} m_{ij}}{m_i}.$$

Подставив последнее выражение в (1.13), приходим к

$$\lambda_{X_1} = \sum_{(i,j)} \lambda_{ij} \pi_i P_{ij} \frac{m_{ij}}{m_i}. \tag{1.14}$$

Рассмотрим семейство $\varkappa := \{\{\varkappa_{ij}^{(n)}\}, n \geqslant 0; i, j \geqslant 0\}$ независимых последовательностей (при фиксированных i, j последовательность по n), состоящих из независимых одинаково распределённых случайных величин, принимающих значения в $\mathbb{Z}_+$. Математические ожидания равномерно ограничены $c_{ij} := \mathsf{E}\varkappa_{ij}^{(n)} < C < +\infty$. Предполагается также, что семейство $\varkappa$ не зависит от $R(t)$.

Определение 15. Случайный поток

$$X_2(t) := \sum_{j=1}^{N(t)} \varkappa_{\zeta_{j-1}\zeta_j}^{(j)}, \tag{1.15}$$

называется *полумарковским потоком поступлений* (ПМПП) (semi-Markov arrival process) с управляющим процессом $R(t)$.

Имеем $X_2(0) = 0$, $X_2(t) \geqslant 0$. Интенсивность потока $X_2(t)$ [32](S. Asmussen, 1999) имеет вид

$$\lambda_{X_2} = \sum_{i=0}^{+\infty} \frac{\pi_i}{m_i} \sum_{j=0}^{+\infty} c_{ij} P_{ij}. \tag{1.16}$$

Определение 16. Случайный поток

$$X(t) := X_1(t) + X_2(t),$$

где $X_1(t)$ и $X_2(t)$ заданы соответственно (1.11) и (1.15) и образованы взаимно независимыми семействами Z и $\varkappa$, называется *полумарковским потоком* (semi-Markov flow) с управляющим процессом $R(t)$.

Интенсивность ПМП $X(t)$

$$\lambda_X = \sum_{i=0}^{+\infty} \frac{\pi_i}{m_i} \sum_{j=0}^{+\infty} (\lambda_{ij} m_{ij} + c_{ij}) P_{ij}. \tag{1.17}$$

ПМП является регенерирующим. В качестве точек регенерации выбираются моменты достижения управляющим процессом фиксированного состояния, например, 1. А именно,

$$\theta_k = \inf\{t_n \colon t_n > \theta_{k-1}, \widehat{R}(t_n) = 1\}.$$

Рассмотренные примеры показывают, что класс регенерирующих потоков весьма широк. Он включает в себя большинство потоков, обычно используемых в теории массового обслуживания в качестве процессов, описывающих поступление требований в систему. Основные характеристики регенерирующего процесса описываются в терминах стационарного распределения управляющего процесса (цепи Маркова или полумарковского процесса), интенсивностей поступления требований в течение времени пребывания этого процесса в фиксированном состоянии.

1.3 Описание модели и основные понятия

Рассматривается одноканальная система массового обслуживания с неограниченной очередью и регенерирующим входящим потоком $X(t)$, определенным на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ с фильтрацией $\{\mathscr{F}_t, t \geqslant 0\}$, к которой он адаптирован. Случайные моменты $\{\theta_i^{(1)},\, i \geqslant 1\}$ $(\mathsf{P}(\theta_0^{(1)} < \infty) = 1)$ — точки регенерации потока, $\tau_i^{(1)} = \theta_i^{(1)} - \theta_{i-1}^{(1)}$ — длина i-го периода регенерации, $\gamma_i^{(1)} = X(\theta_i^{(1)}) - X(\theta_{i-1}^{(1)})$. Предполагается, что $\mathsf{E}\tau_2^{(1)} < \infty$, $\mathsf{E}\gamma_2^{(1)} < \infty$. По свойству 1 интенсивность потока $\lambda_x = \lim\limits_{t\to\infty} \frac{X(t)}{t} = \frac{\mathsf{E}\gamma_2^{(1)}}{\mathsf{E}\tau_2^{(1)}}$.

Здесь процесс $X(t)$ — это объем работы, который поступил в систему за время $[0,t]$.Заметим, что мы не накладываем никаких условий на отдельные времена обслуживания, они могут быть зависимыми в рамках одного периода регенерации.

Пусть $Y(t)$ — количество работы, которое может выполнить система за время $(0,t]$, если бы в системе всегда были требования. Считаем, что процессы $X(t)$ и $Y(t)$ не зависят друг от друга. В классической модели, когда прибор надежен и работает с единичной скоростью $Y(t) = t$. Здесь мы предполагаем, что

$$Y(t) = \int_0^t e(s)ds,$$

где $e(t,w)$ неотрицательный случайный процесс, все траектории которого интегрируемы по Лебегу в $(0,t]$. Кроме того, мы считаем, что процесс $\{e(t),\, t \geqslant 0\}$ — регенерирующий с точками регенерации $\{\theta_i^{(2)}\}_{i=1}^{\infty}$ $(\mathsf{P}(\theta_0^{(2)} < \infty) = 1)$ и случайные величины $\tau_i^{(2)} = \theta_i^{(2)} - \theta_{i-1}^{(2)}$, $\zeta_j = Y(\theta_j^{(2)}) - Y(\theta_{j-1}^{(2)}) = \int\limits_{\theta_{j-1}^{(2)}}^{\theta_j^{(2)}} e(s)ds$ имеют конечные математические ожидания

$$\mathsf{E}\tau_2^{(2)} < \infty, \quad \mathsf{E}\zeta_2 < \infty.$$

Тогда $Y(t)$ будет регенерирующим потоком и его интенсивность $\lambda_y = \frac{\mathsf{E}\zeta_2}{\mathsf{E}\tau_2^{(2)}}$, а *коэффициент загрузки системы*

$$\rho = \frac{\lambda_x}{\lambda_y} = \frac{\mathsf{E}\gamma_2^{(1)}}{\mathsf{E}\zeta_2}\frac{\mathsf{E}\tau_2^{(2)}}{\mathsf{E}\tau_2^{(1)}}.$$

В модели, когда прибор может находиться в рабочем или нерабочем (он ремонтируется) состоянии, процесс

$$e(t) = \begin{cases} 0, & \text{прибор сломан} \\ 1, & \text{иначе.} \end{cases}$$

Существуют различные дисциплины обслуживания требований после прерывания (см. [47](Gaver, 1962)). В данной модели мы считаем, что при поломке прибора происходит немедленное прерывание обслуживания требования. Прерванное обслуживание продолжается после восстановления прибора, причем время обслуживания требования после восстановления прибора совпадает с остаточным временем обслуживания этого требования в момент поломки. Такая дисциплина называется *preemptive-resume interruptions*.

Введем процесс виртуального времени ожидания $W(t)$, представляющий собой суммарное остаточное время обслуживания требований, находящихся в системе в момент t. Как известно (см., например, [68](Такач, 1967),[13](Боровков, 1972)), для данной модели будет справедливо следующее равенство

$$W(t) = \max[W(0) + Z(t), \sup_{0\leqslant s\leqslant t}(Z(t) - Z(s))], \tag{1.18}$$

где $Z(t) = X(t) - Y(t)$. Приведем условие, при котором процесс $Z(t)$ при $t \to \infty$ стремится по распределению к процессу $\tilde{Z}(t)$ со стационарными приращениями.

Условие 2. В непрерывном случае: распределения случайных величин $\tau_2^{(1)}$ и $\tau_2^{(2)}$ удовлетворяют условию 1. В дискретном случае: распределения случайных величин $\tau_2^{(1)}$ и $\tau_2^{(2)}$ апериодичны.

Действительно, по свойству 3 процессы $X(t)$ и $Y(t)$ по распределению стремятся к процессам со стационарными приращениями $\tilde{X}(t)$ и $\tilde{Y}(t)$, а так как они независимы, то это справедливо и для $Z(t)$.

Лемма 1. *Если существует предельное распределение* $\pi(x) = \lim\limits_{t\to\infty} \mathsf{P}(e(t) \leqslant x)$, *то интенсивность потока* $Y(t)$ *равна* $\lambda_y = \int\limits_0^\infty x d\pi(x)$. *Если процесс* $e(t)$ *принимает только два значения* $\{0, 1\}$ *и существует* $\lim\limits_{t\to\infty} \mathsf{P}(e(t) = 1) = \pi$, *то* $\lambda_y = \pi$.

Доказательство. По определению $\lambda_y = \lim\limits_{t\to\infty} \frac{Y(t)}{t}$ с вероятностью 1. С другой стороны, для процесса $Y(t) = \int\limits_0^t e(s)ds$, имеем

$$\lim_{t\to\infty} t^{-1}\mathsf{E}Y(t) = \lim_{t\to\infty} t^{-1}\int_0^t \mathsf{E}e(s)ds = \int_0^\infty x d\pi(x).$$

Таким образом, $\lambda_y = \int\limits_0^\infty x d\pi(x)$. ■

Предельное распределение у случайного процесса $\{e(t),\, t \geqslant 0\}$ существует, если выполнены условия теоремы Смита, в частности, предельное распределение существует при выполнении условия 2.

Следующее условие позволяет доказать наличие общих точек регенерации для потоков $X(t)$ и $Y(t)$.

Условие 3. Для n-го $(n \geqslant 1)$ периода регенерации потока $Y(t)$ имеет место представление:

$$\tau_n^{(2)} = v_n^{(1)} + v_n^{(2)},$$

где $v_n^{(1)}$ и $v_n^{(2)}$ независимы, $\mathsf{P}(v_n^{(1)} > x) = e^{-\delta x}$, $\delta \in (0, \infty)$ и $Y(\theta_{n-1}^{(2)} + v_n^1) = Y(\theta_{n-1}^{(2)})$ с вероятностью 1.

Лемма 2 (О синхронизации)**.** *Пусть выполнено условие 2 в дискретном случае и условие 3 в непрерывном случае. Определим в дискретном случае*

$$T_0 = \min\left\{\theta_j^{(1)} : \bigcup_{l=0}^{\infty}(\theta_j^{(1)} = \theta_l^{(2)})\right\},$$

$$T_n = \min\left\{\theta_j^{(1)} > T_{n-1} : \bigcup_{l=0}^{\infty}(\theta_j^{(1)} = \theta_l^{(2)})\right\},$$

соответственно в непрерывном случае

$$T_0 = \min\left\{\theta_j^{(1)} : \theta_j^{(1)} \in \bigcup_{l=1}^{\infty}(\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)})\right\},$$

$$T_n = \min\left\{\theta_j^{(1)} > T_{n-1} : \theta_j^{(1)} \in \bigcup_{l=1}^{\infty}(\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)})\right\}.$$

Тогда $\{T_n\}_{n=0}^{\infty}$ — *общие моменты регенерации для процессов* $X(t)$, $Y(t)$ *и*

$$\text{в дискретном случае} \qquad \mathsf{E}\tau_2 = \mathsf{E}\tau_2^{(1)}\mathsf{E}\tau_2^{(2)} < \infty,$$
$$\text{в непрерывном случае} \qquad \mathsf{E}\tau_2 = \delta\mathsf{E}\tau_2^{(1)}\mathsf{E}\tau_2^{(2)} < \infty,$$

где $\tau_n = T_n - T_{n-1}$, $n \geqslant 1$.

Доказательство. Первое утверждение леммы следует из определения моментов T_n и свойств экспоненциального распределения. Перейдем ко второй части леммы. Начнем с непрерывного случая. Введем функции

$$h_2(t) = \mathsf{P}\left(\bigcup_{l=1}^{\infty}(t \in (\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)}))\right),$$
$$\mu_2(t) = \min\{l : \theta_l^{(2)} > t\},$$
$$H_2(t) = \sum_{l=0}^{\infty}\mathsf{P}(\theta_l^{(2)} \leqslant t), \quad (H_2(0) = 0).$$

Тогда

$$h_2(t) = \sum_{l=0}^{\infty}\mathsf{P}(t - v_{l+1}^{(1)} < \theta_l^{(2)} \leqslant t) = H_2(t) - \int_0^t H_2(t-y)\delta e^{-\delta y}dy =$$
$$= \int_0^t e^{-\delta(t-y)}dH_2(y).$$

Из основной теоремы восстановления, получаем

$$\lim_{t\to\infty} h_2(t) = \frac{1}{\mathsf{E}\tau_2^{(2)}}\int_0^{\infty} e^{-\delta y}dy = \frac{1}{\delta\mathsf{E}\tau_2^{(2)}}.$$

Теперь положим

$$\nu_0 = \min\{j : \bigcup_{l=1}^{\infty}(\theta_j^{(1)} \in (\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)}))\},$$
$$\nu_k = \min\{j > \nu_{k-1} : \bigcup_{l=1}^{\infty}(\theta_j^{(1)} \in (\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)}))\}.$$

Тогда последовательность $\{\mu_n = \nu_n - \nu_{n-1}\}_{n=2}^{\infty}$ состоит из независимых одинаково распределенных случайных величин и по тождеству Вальда

$$\mathsf{E}\tau_2 = \mathsf{E}\mu_2\mathsf{E}\tau_2^{(1)}.$$

Поэтому, достаточно доказать, что $\mathsf{E}\mu_2 < \infty$.

Положим $h_1(j) = \sum\limits_{k=0}^{\infty} \mathsf{P}(\nu_k = j)$. По теореме Блекуэлла

$$\lim_{j\to\infty} h_1(j) = \begin{cases} (\mathsf{E}\mu_2)^{-1}, & \text{если } \mathsf{E}\mu_2 < \infty \\ 0, & \text{если } \mathsf{E}\mu_2 = \infty. \end{cases}$$

С другой стороны,

$$h_1(j) = \mathsf{P}\left(\bigcup_{l=1}^{\infty}(\theta_j^{(1)} \in (\theta_{l-1}^{(2)}, \theta_{l-1}^{(2)} + v_l^{(1)}))\right) = \mathsf{E}h_2(\theta_j^{(1)}).$$

Так как $\lim\limits_{j\to\infty} \theta_j^{(1)} = \infty$ с вероятностью 1 и $\{\theta_j^{(1)}\}_{j=0}^{\infty}$, $\{\theta_j^{(2)}\}_{j=0}^{\infty}$ независимые последовательности, то по теореме Лебега о мажорирующей сходимости

$$\lim_{j\to\infty} \mathsf{E}h_2(\theta_j^{(1)}) = \frac{1}{\delta \mathsf{E}\tau_2^{(2)}}$$

и, следовательно, $\mathsf{E}\mu_2 = \delta\mathsf{E}\tau_2^{(2)} < \infty$.

В дискретном случае доказательство аналогично, только функция $h_2(j)$ и последовательность $\{\nu_n,\, n \geqslant 0\}$ имеют вид

$$h_2(j) = \mathsf{P}\left(\bigcup_{l=1}^{\infty}(j = \theta_l^{(2)})\right),$$

$$\nu_0 = \min\{j : \bigcup_{l=1}^{\infty}(\theta_j^{(1)} = \theta_{l-1}^{(2)})\},$$

$$\nu_k = \min\{j > \nu_{k-1} : \bigcup_{l=1}^{\infty}(\theta_j^{(1)} = \theta_{l-1}^{(2)})\}.$$

■

1.4 Условия стабильности и нестабильности

В данном разделе мы приведем необходимые и достаточные условия, при которых процесс $W(t)$ будет стохастически ограниченным и эргодичным.

Теорема 1. *Пусть $W(t)$ случайный процесс, определяемый по формуле* (1.18). *Тогда*

1. *Если $\rho > 1$, то $W(t) \xrightarrow[t\to\infty]{} \infty$ с вероятностью 1.*
2. *Если $\rho < 1$ и выполнено условие 2, то существует $\lim\limits_{t\to\infty} \mathsf{P}(W(t) \leqslant x) = \Psi(x)$ и $\Psi(x)$ является функцией распределения, т.е. процесс $W(t)$ эргодичен.*

3. Если $\rho = 1$ и выполнено условие 3, то $W(t) \xrightarrow[t\to\infty]{} \infty$ по вероятности.

Доказательство. Первый пункт теоремы почти сразу следует из равенства (1.18), так как

$$W(t) \geqslant X(t) - Y(t),$$

поэтому

$$\liminf_{t\to\infty} \frac{W(t)}{t} \geqslant \lambda_x - \lambda_y > 0,$$

что и требовалось доказать.

Перейдем ко второму пункту. Для начала покажем, что значение $W(0)$ не влияет на предельное распределение $W(t)$. Действительно, так как $\rho < 1$, то $Z(t) \xrightarrow[t\to\infty]{} -\infty$ с вероятностью 1, следовательно, для любого $\varepsilon > 0$ существует t_ε такое, что при всех $t > t_\varepsilon$

$$\mathsf{P}(Z(t) + W(0) \geqslant 0) < \varepsilon,$$

далее, по формуле (1.18), получаем неравенства при $t > t_\varepsilon$ для любого $x \geqslant 0$

$$\mathsf{P}\left(\sup_{0\leqslant s\leqslant t}(Z(t) - Z(s)) \leqslant x\right) - \varepsilon \leqslant \mathsf{P}(W(t) \leqslant x) \leqslant \mathsf{P}\left(\sup_{0\leqslant s\leqslant t}(Z(t) - Z(s)) \leqslant x\right).$$

Ввиду произвольности ε, имеем для всех $x \geqslant 0$

$$\lim_{t\to\infty} \mathsf{P}(W(t) \leqslant x) = \lim_{t\to\infty} \mathsf{P}\left(\sup_{0\leqslant s\leqslant t}(Z(t) - Z(s)) \leqslant x\right). \qquad (1.19)$$

Как мы замечали ранее, в силу условия 2, имеет место слабая сходимость процесса $Z(t)$ к процессу со стационарными приращениями $\tilde{Z}(t)$. В книге [13, Глава 1, §3](Боровков, 1972) показано, что не ограничивая общности, можно считать, что процесс $Z(t)$ имеет стационарные приращения. Тогда мы имеем следующее равенство по распределению

$$\sup_{0\leqslant s\leqslant t}(Z(t) - Z(s)) = \sup_{0\leqslant u\leqslant t} Z(u).$$

Так как для процесса со стационарными приращениями и отрицательным сносом $\sup_{u\geqslant 0} Z(u) < \infty$ с вероятностью 1, то из (1.19) следует утверждение теоремы.

Докажем третий пункт теоремы. Пусть $\rho = 1$ и выполнено условие 3. Тогда в силу леммы 2, процессы $\{X(t), Y(t)\}$ имеют общие точки регенерации $\{T_n\}_{n=0}^{\infty}$.

Введем вложенный процесс

$$W_n = W(T_n - 0),$$

и обозначим $x_n = X(T_n) - X(T_{n-1})$, $y_n = Y(T_n) - Y(T_{n-1})$, $\zeta_n = T_n - T_{n-1}$ при $n \in \mathbb{N}$. По свойству 1, имеем $\rho = \mathsf{E}x_n/\mathsf{E}y_n$.

Определим вспомогательный случайный процесс W_n^- следующим образом

$$W_0^- = 0, \quad W_{n+1}^- = [W_n^- + x_n - y_n]^+,$$

очевидно, с вероятностью 1 выполнено неравенство при $n \geqslant 0$

$$W_n^- \leqslant W_n. \tag{1.20}$$

Как известно из [25](Феллер, 2010), процесс $W_n^- \stackrel{d}{=} \max\limits_{k \leqslant n} S_k$, где $S_n = \sum\limits_{k=1}^{n}(x_k - y_k)$. Так как $\mathsf{E}(x_k - y_k) = 0$ и $\mathsf{P}(x_k = y_k) \neq 1$, то верхняя грань такого случайного блуждания с вероятностью 1 стремится к $+\infty$, т.е. $\sup\limits_{k \geqslant 0} S_k = \infty$ с вероятностью 1. Следовательно, $W_n^- \xrightarrow[n\to\infty]{} \infty$ по вероятности, и в силу (1.20), $W_n \xrightarrow[n\to\infty]{} \infty$ по вероятности. Пусть $n(t) = \max\{j : T_j < t\}$, покажем, что $W_{n(t)} \xrightarrow[t\to\infty]{} \infty$ по вероятности. С вероятностью 1 справедливо неравенство

$$W_{n(t)} \geqslant W_{n(t)}^- \stackrel{d}{=} \max_{k \leqslant n(t)} S_k. \tag{1.21}$$

Для любых $K > 0$ и $\varepsilon > 0$ существует n_ε такое, что для всех $n \geqslant n_\varepsilon$

$$\mathsf{P}\left(\sup_{l \leqslant n} S_l > K\right) > 1 - \varepsilon.$$

Также существует T_ε такое, что для любого $T > T_\varepsilon$

$$\mathsf{P}(n(T) > n_\varepsilon) > 1 - \varepsilon,$$

следовательно, имеем

$$\mathsf{P}(W_{n(t)}^- > K) \geqslant \mathsf{P}\left(\sup_{l \leqslant n_\varepsilon} S_l > K,\, n(T_\varepsilon) \geqslant n_\varepsilon\right) \geqslant \mathsf{P}\left(\sup_{l \leqslant n_\varepsilon} S_l > K\right) - \varepsilon \geqslant 1 - 2\varepsilon,$$

это вместе с (1.21) означает, что $W_{n(t)} \xrightarrow[t\to\infty]{} \infty$ по вероятности.

Далее, ввиду неравенства

$$W_{n(t)} - y_{n(t)+1} \leqslant W(t),$$

где случайные величины $W_{n(t)}$ и $y_{n(t)+1}$ независимы, процесс $W(t)$ стремится к бесконечности по вероятности. ■

Перенесем этот результат на систему массового обслуживания $Reg/G/1/\infty$ с ненадежным прибором. Для этого введем входящий поток $A(t)$ — количество требований, поступивших в систему за время $(0, t]$. Поток $A(t)$ — регенерирующий с точками регенерации $\{\theta_n^{(1)}\}_{n=0}^{\infty}$. Пусть $\xi_n = A(\theta_n^{(1)}) - A(\theta_{n-1}^{(1)})$

и $\mathsf{E}\xi_n < \infty$, тогда для интенсивности потока $A(t)$ справедливо равенство $\lambda = \frac{\mathsf{E}\xi_n}{\mathsf{E}\tau_n^{(1)}}$.

Времена обслуживания описываются последовательностью $\{\eta_n\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин, $\mathsf{E}\eta_n = b < \infty$. Причем эта последовательность не зависит от процесса $A(t)$. Тогда поток $X(t)$ определяется следующей формулой

$$X(t) = \sum_{k=1}^{A(t)} \eta_k,$$

он тоже будет регенерирующим с точками регенерации $\{\theta_n^{(1)}\}_{n=0}^{\infty}$.

Пусть $\{u_n^{(1)}\}_{n=1}^{\infty}$ и $\{u_n^{(2)}\}_{n=1}^{\infty}$ — последовательности длин интервалов рабочего состояния прибора и ремонта соответственно. В большинстве моделей обслуживания с ненадежным прибором предполагается, что эти последовательности состоят из независимых одинаково распределенных случайных величин. В нашей модели мы не делаем такого ограничения, а требуем чтобы эти интервалы описывались некоторым регенерирующим процессом $e(t)$. Таким образом, наша модель является существенным обобщением в данной области.

Следствие 1. *Теорема 1 верна для системы $Reg/G/1/\infty$ с ненадежным прибором и коэффициент загрузки принимает следующий вид*

$$\rho = \frac{\lambda b}{\pi}.$$

Доказательство. Выразим величины λ_x и λ_y. Для потока $X(t)$, очевидно, получаем с вероятностью 1

$$\lambda_x = \lim_{t\to\infty} \frac{X(t)}{t} = \lim_{t\to\infty} \frac{1}{t} \sum_{n=1}^{A(t)} \eta_n = \lambda b.$$

Как было показано ранее, если выполнено условие 2 или 3, то у процесса $e(t)$ существует предельное стационарное распределение и $\lambda_y = \pi$. ■

Применим следствие 1 к одноканальным системам обслуживания с регенерирующими входящими потоками из раздела 1.2. Предполагаем, что процесс $e(t)$ эргодичен.

Следствие 2. *Пусть входящий поток является*

- *ДСПП и выполнено условие 2, λ выражается по формуле* (1.2)*;*
- *ММП, где λ вычисляется по формуле* (1.6)*;*

- *МПП, где λ вычисляется по формуле* (1.10)*;*
- *потоком со случайной амплитудой (случайным периодом) и выполнено условие 2, λ выражается по формуле* (1.3) *(*(1.4)*);*
- *потоком Льюиса и выполнено условие 2, λ выражается по формуле* (1.9)*;*
- *ПМП и выполнено условие 2, λ выражается по формуле* (1.17)*.*

Тогда при $\lambda b < \pi$ процесс виртуального времени ожидания эргодичен.

1.5 Примеры

Пример 1. Рассмотрим систему обслуживания $Reg/G/1/\infty$ с входящим потоком $A(t)$ и ненадежным прибором. Времена обслуживания образуют последовательность $\{\eta_n\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин с функцией распределения $B(x)$ и средним b.

Прибор может выходить из строя, причем его поломки и интервалы между восстановлениями связаны с функционированием случайного процесса $U(t)$, не зависящего от $A(t)$ и $\{\eta_n\}_{n=1}^{\infty}$. Предполагается, что $U(t)$ — эргодическая цепь Маркова с множеством состояний $\mathbb{E} = (0, 1, 2, \dots)$ и инфинитезимальной матрицей $Q = (q_{ij})$. Такая постановка задачи, где случайная среда определяет работу прибора, является довольно популярной (см., например, [19](Дудин, Назаров, 2015)).

Теперь опишем, как процесс $U(t)$ влияет на прибор. В момент перехода $U(t)$ в состояние $i\,(i \in \mathbb{E})$ работающий прибор выходит из строя с вероятностью $\alpha_i \geqslant 0$, а сломанный прибор восстанавливается с вероятностью $\beta_i \geqslant 0$. Таким образом, поломки и восстановления прибора возникают только в моменты изменений состояния процесса $U(t)$.

Предложенная модель может описывать различные ситуации. Если, например, $U(t)$ — число требований в системе $M/M/1$ и $\alpha_1 = 1$, $\beta_0 = 1$, $\alpha_i \neq 0\,(i \neq 1)$, $\beta_i = 0\,(i \neq 0)$, то мы по сути рассматриваем систему с абсолютным приоритетом. Имеются требования двух типов, $A(t)$ — входящий поток требований первого типа, а требования второго типа поступают в соответствии с пуассоновским потоком и имеют экспоненциально распределенное время обслуживания. Требования второго типа обладают абсолютным приоритетом, так что требование первого типа может обслуживаться, если в системе нет требований второго типа.

Другая ситуация возникает, когда работоспособность прибора зависит от внешних случайных факторов. Множество состояний $U(t)$ представляется в виде объединения двух множеств $E = E_+ \cup E_-$, причем

$$\begin{cases} \alpha_i & > 0 \\ \beta_i & = 0 \end{cases} \text{при}\, i \in E_- \quad \text{и} \quad \begin{cases} \alpha_i & = 0 \\ \beta_i & > 0 \end{cases} \text{при}\, i \in E_+,$$

то есть в состояниях из E_- прибор может сломаться, а в состояниях из E_+ он восстанавливается.

Далее, мы предполагаем, что выполнено следующее условие.

Условие 4. У процесса $U(t)$ найдутся состояния i_0 и i_1, такие что $\alpha_{i_0} > 0$, $\beta_{i_1} > 0$.

Не ограничивая общности, положим $i_0 = 0$.

Поскольку все состояния сообщаются друг с другом, то интервалы, в течение которых прибор находится в рабочем (нерабочем) состоянии, ввиду условия 4, конечны с вероятностью единица.

Если прибор выходит из строя в момент, когда на нем обслуживается требование, то в момент восстановления прибора обслуживание требования продолжается с того момента, где оно было прервано, так что время обслуживания имеет функцию распределения $B(x)$.

Данная модель отличается от рассмотренной в [47] тремя обстоятельствами. В [47] предполагается, что входящий поток пуассоновский, то есть рассматривается система $M/G/1/\infty$, а времена рабочего и нерабочего состояния — независимые случайные величины. Кроме того, считается, что время рабочего состояния имеет экспоненциальное распределение. Также в [47] считается, что поломка прибора может возникнуть, только когда он обслуживает требование. Эти условия, вообще говоря, не выполняются в рассматриваемой модели. Отдельно заметим, что независимость интервалов рабочего и нерабочего состояний будет иметь место, если, например, существуют j_0 и j_1, такие что

$$\alpha_{j_0} > 0, \text{ а } \alpha_j = 0 \text{ при } j \neq j_0,$$
$$\beta_{j_1} > 0, \text{ а } \beta_j = 0 \text{ при } j \neq j_1,$$

то есть поломка (восстановление) прибора может возникнуть только при одном состоянии $U(t)$.

Определим случайную среду для системы с помощью процесса $N(t) = \{e(t), U(t)\}$, где $e(t) = 1$, если в момент t прибор находится в рабочем состоянии, и $e(t) = 0$ в противном случае. Очевидно, что $N(t)$ также является цепью Маркова с непрерывным временем с множеством состояний $\{0, 1\} \times \mathbb{E}$ и элементы её инфинитезимальной матрицы выглядят следующим образом (при $i \neq j$):

$$\begin{aligned} &q_{(0,i)(0,j)} = q_{ij}(1-\beta_j), && q_{(0,i)(1,j)} = q_{ij}\beta_j, \\ &q_{(1,i)(0,j)} = q_{ij}\alpha_j, && q_{(1,i)(1,j)} = q_{ij}(1-\alpha_j), \\ &q_{(k,i)(k,i)} = q_{ii}, && q_{(k,i)(1-k,i)} = 0. \end{aligned} \tag{1.22}$$

Эргодичность цепи $U(t)$ влечет за собой эргодичность $N(t)$. Положим

$$\begin{aligned} p_i^0 &= \lim_{t\to\infty} \mathsf{P}(e(t)=0, U(t)=i), \\ p_i^1 &= \lim_{t\to\infty} \mathsf{P}(e(t)=1, U(t)=i), \\ \pi &= \lim_{t\to\infty} \mathsf{P}(e(t)=1) = \sum_{i\in\mathbb{E}} p_i^1. \end{aligned} \tag{1.23}$$

Стационарные вероятности p_i^0, p_i^1 являются решением системы

$$p_i^0 = \sum_{j\in\mathbb{E}} \left(p_j^0 p_{(0,j)(0,i)} + p_j^1 p_{(1,j)(0,i)}\right), \quad p_i^1 = \sum_{j\in\mathbb{E}} \left(p_j^0 p_{(0,j)(1,i)} + p_j^1 p_{(1,j)(1,i)}\right),$$

где

$$p_{(k,j)(l,i)} = \frac{q_{(k,j)(l,i)}}{-q_{(k,j)(k,j)}}, \quad k,l \in \{0,1\}, \quad i,j \in \mathbb{E},$$

а $q_{(k,j)(l,i)}$ находятся из (1.22).

Введем процесс виртуального времени ожидания $W(t)$ и обозначим коэффициент загрузки $\rho = \frac{\lambda b}{\pi}$. Тогда справедлив следующий результат.

Следствие 3. *Если $U(t)$ — эргодическая цепь Маркова и выполнено условие 4, то утверждения следствия 1 верны для этой системы.*

Доказательство. Достаточно показать, что процесс $e(t)$ удовлетворяет условию 3, из этого будет следовать выполнение условия 2. Зафиксируем состояние $(0,0)$ цепи Маркова $N(t)$ и определим следующие случайные величины

$$T_j = \min\{t > T_{j-1} : N(t) = (0,0)\}, \quad T_0 = 0,$$

то есть $\{T_n\}_{n=0}^{\infty}$ — это последовательные моменты возвращения процесса $N(t)$ в состояние $(0,0)$. Тогда $\{T_n\}_{n=0}^{\infty}$ являются точками регенерации процесса $N(t)$, а, следовательно, и процесса $e(t)$. Более того, в силу условия 4 и эргодичности цепи $N(t)$ мы получаем, что $\mathsf{E}(T_n - T_{n-1}) < \infty$. А так как $N(t)$ это цепь Маркова, то период регенерации $T_n - T_{n-1}$ представим в виде суммы

$$T_n - T_{n-1} = v_n^{(1)} + v_n^{(2)},$$

где $v_n^{(1)}$ — время сидения процесса $N(t)$ в состоянии $(0,0)$, то есть оно имеет экспоненциальное распределение с параметром $-q_{(0,0)(0,0)}$. Таким образом, процесс $e(t)$ удовлетворяет условию 3 и мы можем применить следствие 1. ■

Пример 2. В качестве примера применения следствия 3 рассмотрим упоминавшуюся ранее систему с приоритетами. В одноканальную систему массового обслуживания поступают требования двух типов, причем требования второго типа имеют абсолютный приоритет по отношению к требованиям первого типа. Входящие потоки $A_i(t), i = 1, 2$ независимы и $A_1(t)$ — регенерирующий с интенсивностью λ_1, а $A_2(t)$ — пуассоновский с параметром λ_2. Времена обслуживания требований первого типа — независимые одинаково распределенные случайные величины с функцией распределения $B(x)$ и средним b, а времена обслуживания требований второго типа распределены экспоненциально с параметром μ_2. Требования первого типа обслуживаются, когда в системе нет требований второго типа, а прерванное обслуживание требования первого типа продолжается после ухода из системы всех требований второго типа.

В качестве $U(t)$ рассмотрим число требований второго типа в момент t. Это — цепь Маркова и в предположении $\rho_2 = \frac{\lambda_2}{\mu_2} < 1$ у неё существует стационарное распределение, при этом $R_j = \lim_{t\to\infty} \mathsf{P}(U(t) = j) = (1 - \rho_2)\rho_2^j$. По отношению к требованиям первого типа прибор прерывает обслуживание, когда в системе появляются требования второго типа, то есть $\alpha_0 = 0, \beta_0 = 1, \alpha_i = 1, \beta_i = 0$ при$i \geqslant 1$, так что $\pi = R_0$ (см., (1.23)) и коэффициент загрузки имеет вид $\rho_1 = \frac{\lambda_1 b}{1-\rho_2}$. Если $A_1(t)$ —пуассоновский процесс, то получаем известные условия эргодичности (см., например,[23](Саати, 1971)).

Пример 3. В этом примере рассмотрим систему $Reg/G/1/\infty$ функционирующую в случайной среде $U(t)$, здесь $U(t)$ — число требований в системе $M/M/1/\infty$ с пуассоновским входящим потоком интенсивности λ_1 и экспоненциально распределенным временем обслуживания с параметром μ_1. Как известно из [23](Саати, 1971), цепь $U(t)$ эргодическая, если $\lambda_1 < \mu_1$. Инфинитезимальная матрица для процесса $U(t)$ выглядит следующим образом

$$\begin{aligned} q_{ii+1} &= \lambda_1,\ i = 0, 1, \dots, \\ q_{ii-1} &= \mu_1,\ i = 1, 2, \dots, \\ q_{ij} &= 0,\ |i - j| > 1. \end{aligned} \tag{1.24}$$

При переходе $U(t)$ в любое состояние работающий прибор выходит из строя с вероятностью α, а сломанный восстанавливается с вероятностью β. То есть $\alpha_i = \alpha$, $\beta_i = \beta$ при $i \in \{0, 1, \dots\}$. Стоит отметить, что в данной системе интервалы рабочего и нерабочего состояния будут зависимы.

Введем обозначения

$$P_j = \lim_{t\to\infty} \mathsf{P}(U(t) = j, e(t) = 1),\ P(z) = \sum_{j=1}^{\infty} z^j P_j,$$

$$Q_j = \lim_{t\to\infty} \mathsf{P}(U(t) = j, e(t) = 0),\ Q(z) = \sum_{j=1}^{\infty} z^j Q_j,$$

$$R_j = P_j + Q_j,\ R(z) = \sum_{j=1}^{\infty} z^j R_j,\ \rho_1 = \frac{\lambda_1}{\mu_1}.$$

Теперь, для того чтобы найти условие эргодичности данной системы, нужно найти $\pi = \lim\limits_{t\to\infty} \mathsf{P}(e(t) = 1) = \sum\limits_{j=0}^{\infty} P_j$.

Для этого воспользуемся тем, что $R_j = \lim\limits_{t\to\infty} \mathsf{P}(U(t) = j) = (1-\rho_1)\rho_1^j$ и выпишем систему уравнений, используя (1.24).

$$\begin{aligned} \lambda_1 P_0 &= (1-\alpha)\mu_1 P_1 + \beta\mu_1 Q_1, \\ (\lambda_1+\mu_1)P_j &= (1-\alpha)\lambda_1 P_{j-1} + (1-\alpha)\mu_1 P_{j+1} + \beta\lambda_1 Q_{j-1} + \beta\mu_1 Q_{j+1}, j = 1, 2, \ldots. \end{aligned}$$

Полагая $c = 1-\alpha-\beta$, очевидным образом получаем соотношение для производящей функции $P(z)$

$$P(z)(-c\rho_1 z^2 + (1+\rho_1)z - c) = \frac{1-\rho_1}{1-\rho_1 z}\beta(1+\rho_1 z^2) + P_0(z-c) - \beta(1-\rho_1). \quad (1.25)$$

Отсюда

$$P(z) = \frac{(1-\rho_1)\beta(1+\rho_1 z^2) + (1-\rho_1 z)(P_0(z-c) - \beta(1-\rho_1))}{(-c\rho_1 z^2 + (1+\rho_1)z - c)(1-\rho_1 z)}.$$

Чтобы выразить P_0, находим корни

$$z_{1,2} = \frac{(1+\rho_1) \pm \sqrt{(1+\rho_1)^2 - 4c^2\rho_1}}{2c\rho_1}$$

уравнения $g(z) = z^2 - \frac{1+\rho_1}{\rho_1 c} z + \frac{1}{\rho_1} = 0$.

Из этих двух корней выбираем z_2, так как $z_2 \in (-1, 1)$ и, следовательно, $P(z)$ аналитична в этой точке. Используя это обстоятельство, из (1.25) получаем

$$P_0 = \frac{\beta\rho_1(1-\rho_1)z_2(1+z_2)}{(c-z_2)(1-z_2\rho_1)},$$

$$P(1) = \frac{P_0}{1+\rho_1} + \frac{2\beta\rho_1}{(1-c)(1+\rho_1)}.$$

А так как $P(1) = \sum\limits_{j=0}^{\infty} P_j$, то коэффициент загрузки данной системы равен

$$\rho = \frac{\lambda b(\alpha+\beta)(1+\rho_1)}{(\alpha+\beta)P_0 + 2\beta\rho_1},$$

где λ — интенсивность входящего потока, а b — математическое ожидание времени обслуживания. Теперь рассмотрим некоторые предельные случаи:
1) Пусть $c = 1 - h_\alpha - h_\beta$, $\alpha = h_\alpha$, $\beta = h_\beta$, где $h_\alpha > 0$, $h_\beta > 0$ и стремятся к нулю. В таких условиях получаем

$$P(1) \sim \frac{h_\beta}{h_\alpha + h_\beta} \frac{1 + 2\rho_1 - \rho_1^2}{(1+\rho_1)(2-\rho_1)}.$$

Тогда, если $h_\alpha \to 0$ и $h_\beta = o(h_\alpha)$, то $P(1) \to 0$,
если $h_\beta \to 0$ и $h_\alpha = o(h_\beta)$, то $P(1) \to \frac{1+2\rho_1-\rho_1^2}{(1+\rho_1)(2-\rho_1)}$,
если $h_\beta \to 0$ и $h_\alpha = lh_\beta$, то $P(1) \to \frac{1}{1+l}\frac{1+2\rho_1-\rho_1^2}{(1+\rho_1)(2-\rho_1)}$.
2) Пусть $c = -1 + h_\alpha + h_\beta$, $\alpha = 1 - h_\alpha$, $\beta = 1 - h_\beta$, где $h_\alpha > 0$, $h_\beta > 0$ стремятся к нулю. Тогда

$$P(1) \sim \frac{1-\rho_1}{2(1+\rho_1)} + \frac{2\rho_1(1-h_\beta)}{(1+\rho_1)(2-h_\alpha-h_\beta)} \xrightarrow[h_\alpha + h_\beta \to 0]{} \frac{1}{2}.$$

3)Пусть $c = h, 1 - \alpha = \beta + h$. Тогда, если $|h| \to 0$, то $P(1) \to \beta$. Заметим, что если устремить α к нулю, а β к единице, то мы получим модель с надежным прибором. Для такой модели $\rho = \lambda b$.

Пример 4. В следующем примере не удается построить марковский процесс $U(t)$, определяющий факт доступности прибора для обслуживания. Однако, в вспомогательной модели, которая функционирует в марковской случайной среде и мажорирует исходную систему, можно найти достаточное условие эргодичности. Оно близко к необходимому в том смысле, что если система неэргодична, то условие нарушается.

Модель состоит из двух последовательно расположенных систем S_1 и S_2.

Здесь S_1 — одноканальная система с пуассоновским входящим потоком с параметром λ_1. Время обслуживания имеет экспоненциальное распределение с параметром ν_1 (то есть S_1 это $M/M/1/\infty$).

Система S_2 — одноканальная система с двумя входящими потоками. Первый поток требований — это те требования, которые поступают в очередь S_2 из системы S_1. Второй поток требований является пуассоновским с параметром λ_2. Предполагается, что требование из второго потока поступает в

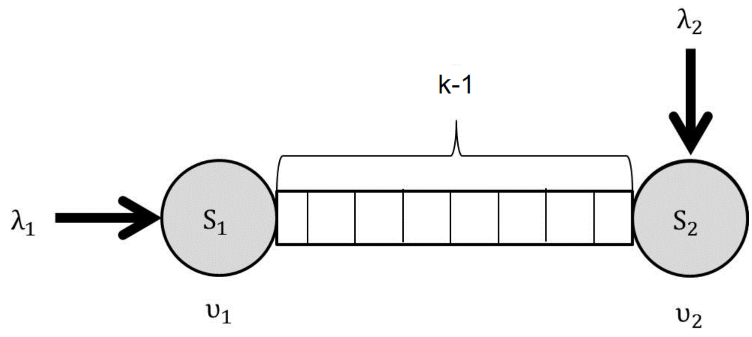

Рис. 1.1:

систему S_2, если прибор не занят и требование получает отказ, если прибор занят. Время обслуживания в системе S_2 распределено экспоненциально с параметром ν_2.

Между S_1 и S_2 есть буфер из $k-1(k<\infty)$ мест. Так что S_1 перестает обслуживать требования, если буфер полностью занят, то есть в системе S_2 есть k требований.

Пусть $q_i(t)$ — число требований в S_i в момент $t(i=1,2)$. В силу сделанных предположений $Q(t)=(q_1(t),q_2(t))$ — неприводимая цепь Маркова, и поскольку $q_2(t)\leqslant k<\infty$, она эргодична тогда и только тогда, когда процесс $q_1(t)$ стохастически ограничен [8].

Рассмотрим вспомогательную систему $(\tilde{S}_1,\tilde{S}_2)$, предположив, что в $\tilde{S}_1$ есть запасные требования, которые поступают на обслуживание, когда прибор освобождается от обслуживания основных, то есть тех, которые поступают с интенсивностью λ_1. Если в момент поступления основного требования прибор занят обслуживанием запасного, он прекращает его обслуживание и начинает обслуживать основное. Поскольку мы считаем, что запасных требований сколь угодно много, поток поступающих в $\tilde{S}_2$ требований будет пуассоновским с параметром ν_2. Разумеется, возникают задержки, если буфер заполнен.

Пусть $\tilde{q}_1(t)$ — число требований в $\tilde{S}_1$ в момент t, а $\tilde{q}_2(t)$ — процесс $q_2(t)$ для $\tilde{S}_2$. Тогда $\tilde{q}_2(t)$ является цепью Маркова, а при одинаковых начальных условиях выполняется стохастическое неравенство

$$q_1(t)\leqslant\tilde{q}_1(t)\quad\text{при}\,t\geqslant 0. \tag{1.26}$$

Мы можем рассматривать $\tilde{q}_2(t)$ как случайную среду для системы $\tilde{S}_1$. При этом
$\alpha_0=\cdots=\alpha_{k-1}=0,\alpha_k=1;\ \beta_0=\cdots=\beta_{k-1}=1,\ \beta_k=0.$

Мы заключаем, что в $\tilde{S}_1$ прибор не работает, если $\tilde{q}_2(t)=k$. По следствию 3, коэффициент загрузки для $\tilde{S}_1$ определяется соотношением

$$\tilde{\rho}=\frac{\lambda_1}{\nu_1(1-\tilde{R}_k)},$$

где $\tilde{R}_k = \lim_{t\to\infty} \mathsf{P}(\tilde{q}_2(t) = k)$.
Для стационарного распределения $\tilde{R}_j = \lim_{t\to\infty} \mathsf{P}(\tilde{q}_2(t) = j)$ имеем систему уравнений

$$
\begin{aligned}
(\nu_1 + \lambda_2)\tilde{R}_0 &= \nu_2 \tilde{R}_1,\\
(\nu_1 + \nu_2)\tilde{R}_1 &= (\nu_1 + \lambda_2)\tilde{R}_0 + \nu_2 \tilde{R}_2,\\
(\nu_1 + \nu_2)\tilde{R}_j &= \nu_1 \tilde{R}_{j-1} + \nu_2 \tilde{R}_{j+1}, \quad 1 < j < k\\
\nu_2 \tilde{R}_k &= \nu_1 \tilde{R}_{k-1}.
\end{aligned}
$$

Отсюда находим, что $\tilde{R}_k = \dfrac{\nu_1^{k-1}(\nu_1 + \lambda_2)(\nu_2 - \nu_1)}{\nu_2^k(\nu_2 - \nu_1) + (\nu_1 + \lambda_2)(\nu_2^k - \nu_1^k)}$.
И таким образом, коэффициент загрузки равен

$$\tilde{\rho} = \frac{\lambda_1}{\nu_1 \nu_2} \cdot \frac{\nu_2^{k+1} - \nu_1^{k+1} + \lambda_2(\nu_2^k - \nu_1^k)}{\nu_2^k - \nu_1^k + \lambda_2(\nu_2^{k-1} - \nu_1^{k-1})}.$$

Если $\nu_1 = \nu_2$, то $\tilde{R}_k = \frac{\nu_1 + \lambda_2}{\nu_1(k+1) + k\lambda_2}$ и коэффициент загрузки равен

$$\tilde{\rho} = \frac{\lambda_1}{\nu_1} \cdot \frac{\nu_1(k+1) + k\lambda_2}{k\nu_1 + \lambda_2(k-1)}.$$

Пусть $\tilde{\rho} < 1$, то есть процесс $\tilde{q}_1(t)$ эргодичен, а стало быть, стохастически ограничен. Из (1.26) получаем стохастическую ограниченность $q_1(t)$, а, следовательно, и эргодичность этого процесса.

Предположим, что $Q(t)$ не эргодичен. Тогда $q_1(t)$ стохастически неограничен, а в силу (1.26) стохастически неограничен $\tilde{q}_1(t)$. Это означает, что $\tilde{\rho} \geqslant 1$.

Следствие 4. *Если $\tilde{\rho} < 1$, то $Q(t)$ эргодичен. Если $Q(t)$ неэргодичен, то $\tilde{\rho} \geqslant 1$.*

В этой главе рассмотрена одноканальная система обслуживания с регенерирующим потоком и прибором, скорость работы которого определяется регенерирующим процессом. Такая модель включает в себя и случай ненадежного прибора, когда скорость работы на некоторых интервалах времени равна нулю. Получены необходимые и достаточные условия эргодичности виртуального времени ожидания. Одним из важных обобщений модели с ненадежным прибором является то, что не предполагается независимость интервалов рабочего состояния прибора и времени ремонта. В дальнейшем, подход, предложенный в главе, может быть применен к многоканальным системам с регенерирующим входящим потоком и приборами, у которых моменты поломок и восстановления определяются регенерирующими процессами.

Глава 2

Система $M/G/1/\infty$ с приоритетным обслуживанием и ненадежным прибором

В этой главе рассмотрим несколько моделей типа $M/G/1/\infty$ с ненадежным прибором и приоритетным обслуживанием. Для них получены предельные распределения процесса виртуального времени ожидания и числа требований в системе в терминах преобразования Лапласа-Стилтьеса.

Для этих систем мы найдем асимптотику вероятностей больших уклонений процесса виртуального времени ожидания в стационарном режиме.

2.1 Предыдущие результаты по теме

Приведем некоторые результаты из [18](Бочаров, Печинкин, 1995) и [47](Gaver, 1962), которые будут использованы в дальнейшем.

Одной из первых работ, где изучались системы типа $M/G/1/\infty$ с выходами прибора из строя и последующими восстановлениями, является статья [47] (Gaver, 1962). В этой работе рассматривается вариант, когда поломка прибора может происходит только во время обслуживания требования. Можно сделать различные предположения о том, что происходит, когда прибор выходит из строя в процессе обслуживания требования. Выделили следующие две наиболее распространенные дисциплины (см. [47]):

D1 (*Preemptive-resume interruptions*). Прерывание обслуживания интерпретируется как поломка прибора и, соответственно, имеет немедленный эффект. Прерванное обслуживание продолжается после восстановления прибора, причем время обслуживания требования после восстановления прибора совпадает с остаточным временем обслуживания этого требования в момент поломки.

D2 (*Preemptive-repeat-different interruptions*). Обслуживание требования прерывается немедленно, а время обслуживания требования после восстановления прибора имеет то же распределение, что и начальное время обслуживания, и не зависит от него.

Две эти дисциплины были введены в работе [47](Gaver, 1962), там же можно найти описание и других дисциплин обслуживания требования после поломки прибора. Также в [47] было введено понятие полного времени обслуживания (completion time).

Определение 17. *Полное время обслуживания* требования - это время, которое требование находится на приборе вне зависимости от того, исправен прибор или нет.

Приведем описание системы и некоторые результаты, полученные в [18, Глава 7](Бочаров, Печинкин, 1995). В этой книге рассмотрена система, где предполагается, что прибор может выходить из строя не только во время обслуживания требования, но и когда требований в системе нет, что является обобщением системы из [47].

Рассматривается система массового обслуживания $M/G/1/\infty$ с ненадежным прибором. Вероятность выхода прибора из строя на интервале времени $(t, t+\Delta)$ зависит только от состояния системы в момент t и, если система в момент t свободна, то равна $\mu^*\Delta + o(\Delta)$, а если на приборе находится требование, то равна $\mu\Delta + o(\Delta)$. Параметры μ^* и μ — интенсивности отказов в свободном и занятом состояниях соответственно.

Если в момент отказа прибора система свободна, то прибор ремонтируется случайное время, имеющее функцию распределения $G^*(x)$ с преобразованием Лапласа-Стилтьеса (ПЛС) $\zeta^*(s)$ и средним g^*. При этом первое требование, поступившее в свободную систему в момент ремонта прибора, становится на прибор, но его обслуживание начинается только после окончания ремонта. Остальные поступающие требования скапливаются в очереди.

Если же в момент отказа прибора на нем находится требование, то обслуживание прекращается, а прибор ремонтируется случайное время с функцией распределения $G(x)$, ПЛС $\zeta(s)$ и средним g. В течение этого времени требование продолжает находиться на приборе, причем после восстановления прибора требование дообслуживается остаточное время.

Время обслуживания требования имеет функцию распределения $B(x)$, математическое ожидание b и ПЛС $\beta(s)$.

Пусть $\alpha(s)$ и $\alpha^*(s)$ — ПЛС времени пребывания требования на приборе (полное время обслуживания), пришедшего в занятую систему и свободную соответственно. В [47] получены формулы, позволяющие выразить $\alpha(s)$ через исходные данные для дисциплин D1 и D2:

$$\text{D1} \qquad \alpha(s) = \beta(s + \mu - \mu\zeta(s)), \tag{2.1}$$

$$\text{D2} \qquad \alpha(s) = \frac{\beta(s+\mu)}{1 - \zeta(s)\frac{\mu}{s+\mu}(1 - \beta(s+\mu))}, \tag{2.2}$$

где λ — интенсивность поступления требований в систему. Формулу для $\alpha^*(s)$ можно найти в [18], она имеет следующий вид

$$\alpha^*(s) = \frac{\lambda}{\lambda + \mu^* - \mu^*\zeta^*(\lambda)} \left[1 + \mu^* \frac{\zeta^*(s) - \zeta^*(\lambda)}{\lambda - s}\right] \alpha(s). \tag{2.3}$$

Зная $\alpha(s)$ и $\alpha^*(s)$, получаем соответствующие им математические ожидания:

$$\text{D1} \qquad a = -\alpha'(0) = b(1 + \mu g), \tag{2.4}$$

$$\text{D2} \qquad a = \frac{1 - \beta(\mu)}{\beta(\mu)} \left(d + \frac{1}{\mu}\right), \tag{2.5}$$

$$a^* = \frac{\mu^*}{\lambda + \mu^* - \mu^*\zeta^*(\lambda)} \left[g^* - \frac{1 - \zeta^*(\lambda)}{\lambda}\right] + a. \tag{2.6}$$

Введем процесс $q(t)$ — число требований в системе в момент t. Для определенности будем считать, что в начальный момент 0 система свободна и прибор находится в исправном состоянии. Тогда $q(0) = 0$.

Пусть $\tau_0 = 0$, а $\{\tau_n\}_{n=1}^{\infty}$ — последовательность моментов ухода требований из системы. Так как после каждого ухода требования прибор обязательно находится в исправном состоянии, то $\{q_n = q(\tau_n + 0), n \geqslant 0\}$ — однородная цепь Маркова.

Для этой цепи Маркова в [18] доказана следующая теорема.

Теорема (Бочаров, Печинкин). *Если $\rho = \lambda a < 1$, то существуют предельные (стационарные) вероятности состояний $p_i = \lim\limits_{n\to\infty} \mathsf{P}(q_n = i), i = 0, 1, \ldots,$ и для функции $P(z) = \sum_{i=0}^{\infty} p_i z^i$ справедливо следующее соотношение:*

$$P(z) = \frac{\alpha(\lambda - \lambda z) - z\alpha^*(\lambda - \lambda z)}{\alpha(\lambda - \lambda z) - z} p_0,$$

где $p_0 = \frac{1-\rho}{1-\rho+\lambda a^*}$; функции $\alpha(s)$ и $\alpha^*(s)$ определены формулами (2.1), (2.2) *и* (2.3)*; a и a^* — математические ожидания, определенные формулами* (2.4), (2.5) *и* (2.6)*.*

Пусть $W(t)$ — время, через которое после момента t прибор будет способен приступить к обслуживанию нового требования, т.е. он освободится от обслуживания всех требований, находившихся в системе в момент t, и будет в рабочем состоянии.

Если $\rho = \lambda a < 1$, то существует $\lim_{t\to\infty} \mathsf{P}(W(t) \leqslant x) = \Psi(x)$, где $\Psi(x)$ — функция распределения. Из [18, Глава 7] известно, что для функции $w(s) = \int_0^\infty e^{-sx} d\Phi(x)$ справедлива следующая формула

$$w(s) = \frac{s + \mu^* - \mu^*\zeta^*(s)}{s - \lambda + \lambda\alpha(s)} \frac{1-\rho}{1+\mu^* g^*}. \tag{2.7}$$

2.2 Вероятности больших уклонений

В данном разделе мы оценим вероятность образования очень большой очереди для системы, описанной в разделе 2.1. Предполагаем, что обслуживание после поломки соответствует дисциплине D1. Задача больших уклонений с дисциплиной D2 будет рассмотрена как пример в следующей главе для более общей модели. Наша цель — исследовать асимптотику $1-\Psi(x)$ при $x \to \infty$, в предположении, что время обслуживания и время ремонта имеют "тяжелые хвосты".

Далее для любой функции распределения $F(x)$ с ПЛС $\varphi(s)$ и n конечными моментами f_i $(i = \overline{1,n})$ будем обозначать

$$\varphi_n(s) := (-1)^{n+1} \left\{ \varphi(s) - \sum_{k=0}^{n} \frac{f_k}{k!}(-s)^k \right\}, \tag{2.8}$$

то есть для функции $\varphi(s)$ справедливо следующее разложение

$$\varphi(s) = \sum_{k=0}^{n} \frac{f_k}{k!}(-s)^k + (-1)^{n+1}\varphi_n(s). \tag{2.9}$$

Наш анализ будет опираться на формулу (2.7) и результат, установленный в [36](Bingham, Doney, 1974). Для удобства приведем его в виде теоремы.

Теорема. *(Bingham& Doney [36]) Если $f_i < \infty$ при $i = 1, \dots n$ и $\nu = n + \mu$, где $0 < \mu < 1$, тогда следующие утверждения эквивалентны*

$$\varphi_n(s) \sim s^\nu L(1/s), \text{ при } (s \to 0), \tag{2.10}$$

$$1 - F(x) \sim \frac{(-1)^n}{\Gamma(1-\nu)} x^{-\nu} L(x), \text{ при } (x \to \infty), \tag{2.11}$$

где $f(x) \sim g(x)$ означает, что $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$, а $L(x)$ — медленно меняющаяся функция, т.е. $\lim_{x\to\infty} \frac{L(xt)}{L(x)} = 1$ для любого $t > 0$.

Обозначим моменты для распределений $B(x)$, $G(x)$, $G^*(x)$:

$$b_i := (-1)^i \beta^{(i)}(0), \quad g_i := (-1)^i \zeta^{(i)}(0), \quad g_i^* := (-1)^i \zeta^{*(i)}(0).$$

Предположение о том, что распределения времени обслуживания и времен ремонта имеют "тяжелые хвосты"в данной статье эквивалентно следующему условию.

Условие 5. Пусть l, m и m^* целые числа, большие 1, и при $x \to \infty$

$$\begin{aligned}
1 - B(x) &\sim \frac{(-1)^l}{\Gamma(1-p)} x^{-p} L_1(x), \quad l < p < l+1, \\
1 - G(x) &\sim \frac{(-1)^m}{\Gamma(1-r)} x^{-r} L_2(x), \quad m < r < m+1, \\
1 - G^*(x) &\sim \frac{(-1)^{m^*}}{\Gamma(1-r^*)} x^{-r^*} L_2^*(x), \quad m^* < r^* < m^*+1,
\end{aligned}$$

где $L_1(x), L_2(x)$ и $L_2^*(x)$ — медленно меняющиеся функции.

Определим функции $\beta_l(s)$, $\zeta_m(s)$, ${\zeta^*}_{m^*}(s)$ по формуле (2.8) с заменой функции $\varphi(s)$ на $\beta(s)$, $\zeta(s)$, $\zeta^*(s)$ соответственно. В силу теоремы Bingham& Doney, при $s \to 0$

$$\beta_l(s) \sim s^p L_1(1/s), \quad l < p < l+1, \tag{2.12}$$

$$\zeta_m(s) \sim s^r L_2(1/s), \quad m < r < m+1, \tag{2.13}$$

$$\zeta^*_{m^*}(s) \sim s^{r^*} L_2^*(1/s), \quad m^* < r^* < m^*+1. \tag{2.14}$$

Положим

$$\tilde{n} = \min(l, m, m^*),\ \tilde{q} = \min(p, r, r^*),\ n = \min(l, m),\ q = \min(p, r) \tag{2.15}$$

и определим функции

$$\begin{aligned}
L_3(x) &:= 1(p \leqslant r)(1 + \mu g_1)^p L_1(x) + 1(p \geqslant r) \mu b_1 L_2(x), \\
L_4(x) &:= 1(q \leqslant r^*) \frac{\lambda}{1-\rho} L_3(x) + 1(q \geqslant r^*) \frac{\mu^*}{1 + \mu^* g_1^*} L_2^*(x),
\end{aligned}$$

где $1(A)$ — индикатор события A.

Теорема 2. *Если выполнено условие 5, то при $x \to \infty$*

$$1 - \Psi(x) \sim \frac{(-1)^{\tilde{n}-1}}{\Gamma(2-\tilde{q})} x^{-\tilde{q}+1} L_4(x),$$

где $\tilde{n}$ и $\tilde{q}$ определены в (2.15).

Доказательство теоремы следует из двух лемм и теоремы Bingham& Doney. Пусть $\alpha_n(s)$ определяется формулой (2.8) с заменой $\varphi(s)$ на $\alpha(s)$.

Лемма 3. *Если выполнено условие 5, то при* $s \to 0$

$$\alpha_n(s) \sim s^q L_3(1/s). \tag{2.16}$$

Доказательство. Воспользуемся формулой (2.1) и разложением (2.9) для функции $\beta(s)$. Тогда

$$\alpha(s) = \beta(s+\mu-\mu\zeta(s)) = \sum_{i=0}^{l} \frac{b_i}{i!}(-s-\mu+\mu\zeta(s))^i + (-1)^{l+1}\beta_l(s+\mu-\mu\zeta(s)),$$

при этом $s+\mu(1-\zeta(s)) = s(1+\mu g_1) - \mu\left(\sum_{j=2}^{m} \frac{g_j}{j!}(-s)^j + (-1)^{m+1}\zeta_m(s)\right)$, т.е.

$$\alpha(s) = \sum_{i=0}^{l} \frac{b_i}{i!}\left(-s(1+\mu g_1) + \mu\left(\sum_{j=2}^{m} \frac{g_j}{j!}(-s)^j + (-1)^{m+1}\zeta_m(s)\right)\right)^i + \\ + (-1)^{l+1}\beta_l(s(1+\mu g_1)+o(s)). \tag{2.17}$$

Если $p < r$, то после раскрытия скобок в (2.17) мы получим

$$\alpha(s) = \sum_{i=0}^{l} \frac{a_i}{i!}(-s)^i + (-1)^{l+1}\beta_l(s(1+\mu g_1)+o(s)) + o(s^p).$$

Из (2.12) следует, что $\alpha_l(s) \sim (1+\mu g_1)^p s^p L_1(1/s)$.

Если $p > r$, то из (2.17) имеем

$$\alpha(s) = \sum_{i=0}^{m} \frac{a_i}{i!}(-s)^i + (-1)^{m+1}\mu b_1\zeta_m(s) + o(s^r).$$

Из (2.13) следует, что $\alpha_l(s) \sim \mu b_1 s^r L_2(1/s)$.

Если $p = r$, то (2.17) приводит к разложению

$$\alpha(s) = \sum_{i=0}^{l} \frac{a_i}{i!}(-s)^i + (-1)^{l+1}(\beta_l(s(1+\mu g_1)+o(s)) + \mu b_1\zeta_m(s)) + o(s^p)$$

и из (2.12) и (2.13) следует, что $\alpha_l(s) \sim s^p((1+\mu g_1)^p L_1(1/s) + \mu b_1 L_2(1/s))$. ∎

Теперь мы можем перейти к анализу функции $w(s)$.

Лемма 4. *Пусть выполнено условие 5, тогда при* $s \to 0$

$$w_{\tilde{n}-1}(s) \sim s^{\tilde{q}-1} L_4(1/s). \tag{2.18}$$

Доказательство. Из формулы (2.7) находим

$$w(s) = \frac{1-\rho}{1+\mu^* g_1^*} \frac{s+\mu^*(1-\zeta^*(s))}{s-\lambda(1-\alpha(s))} = \frac{1-\rho}{1+\mu^* g_1^*} \frac{1+\mu^*\left(\frac{1-\zeta^*(s)}{s}\right)}{1-\lambda\left(\frac{1-\alpha(s)}{s}\right)}.$$

Применим разложение (2.9) для функций $\alpha(s)$ и $\zeta^*(s)$

$$\frac{1-\alpha(s)}{s} = -a_1 - \sum_{i=2}^{n} \frac{a_i}{i!}(-s)^{i-1} - (-1)^{n+1}\frac{\alpha_n(s)}{s},$$

$$\frac{1-\zeta^*(s)}{s} = -g_1^* - \sum_{i=2}^{m^*} \frac{g_i^*}{i!}(-s)^{i-1} - (-1)^{m^*+1}\frac{\zeta^*_{m^*}(s)}{s}.$$

Исходя из этих разложений, получаем

$$w(s) = \frac{1+\frac{\mu^*}{1+\mu^* g_1^*}\left(\sum_{j=2}^{m^*}\frac{g_j^*}{j!}(-s)^{j-1}+(-1)^{m^*}\frac{\zeta^*_{m^*}(s)}{s}\right)}{1-\frac{\lambda}{1-\rho}\left(\sum_{i=2}^{n}\frac{a_i}{i!}(-s)^{i-1}+(-1)^{n}\frac{\alpha_n(s)}{s}\right)} =$$

$$= \left(1+\frac{\mu^*}{1+\mu^* g_1^*}\left(\sum_{j=2}^{m^*}\frac{g_j^*}{j!}(-s)^{j-1}+(-1)^{m^*}\frac{\zeta^*_{m^*}(s)}{s}\right)\right) \times$$

$$\times \left(\sum_{k=0}^{\infty}\left(\frac{\lambda}{1-\rho}\left(\sum_{i=2}^{n}\frac{a_i}{i!}(-s)^{i-1}+(-1)^{n}\frac{\alpha_n(s)}{s}\right)\right)^k\right).$$

Обозначив k-ый момент виртуального времени ожидания v_k. После раскрытия скобок мы получаем

$$w(s) = \sum_{k=0}^{\tilde{n}-1}\frac{v_k}{k!}(-s)^k + (-1)^{\tilde{n}}\left(\frac{1(q\leqslant r^*)\lambda}{1-\rho}\frac{\alpha_n(s)}{s} + \frac{1(q\geqslant r^*)\mu^*}{1+\mu^* g_1^*}\frac{\zeta^*_{m^*}(s)}{s}\right) + o(s^{\tilde{q}-1}),$$

т.е.

$$w_{\tilde{n}-1}(s) \sim s^{\tilde{q}-1}\left(1(q\leqslant r^*)\frac{\lambda}{1-\rho}L_3(1/s) + 1(q\geqslant r^*)\frac{\mu^*}{1+\mu^* g_1^*}L_2^*(1/s)\right).$$

■

Применив теорему Bingham& Doney, находим асимптотику $1-\Psi(x)$ при $x\to\infty$.

Замечание 1. Как следует из теоремы 2, асимптотика $1-\Psi(x)$ определяется самым тяжелым хвостом среди функций распределения $B(x)$, $G(x)$ и $G^*(x)$. Также стоит заметить, что теорема 2 справедлива, если одна или две из функций распределения удовлетворяют условию Крамера. Если же все три функции распределения удовлетворяют условию Крамера, то и $\Psi(x)$ будет удовлетворять условию Крамера (см., например, [27](Abate, 1997)).

В следующих разделах мы рассмотрим две модели, в которых объединены два вида прерывания. Первая причина прерывания — это поломка прибора, а вторая — поступление требования высшего приоритета.

2.3 Описание модели M_1

Рассматривается одноканальная система массового обслуживания с ненадежным прибором и двумя типами поступающих требований. Прибор может выходить из строя и восстанавливаться. Этот процесс определяется двумя независимыми последовательностями, каждая из которых состоит из независимых одинаково распределенных случайных величин. При этом время рабочего состояния распределено экспоненциально с параметром ν, а время восстановления имеет функцию распределения $D(x)$, среднее d и ПЛС $\delta(s)$.

Входящие потоки являются пуассоновскими соответственно с параметрами λ_1 и λ_2. Времена обслуживания требований i-го типа $(i = 1, 2)$ образуют последовательность $\{\eta_n^i\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин с функцией распределения $B_i(x)$, средним b_i и ПЛС $\beta_i(s)$.

В данной модели требования второго типа имеют абсолютный приоритет относительно требований первого типа, т.е. если во время обслуживания требования первого типа приходит требование второго типа, то обслуживание неприоритетного требования останавливается и начинается обслуживание приоритетного.

Модель можно рассматривать с точки зрения требований первого типа (эту систему будем обозначать S_1) и с точки зрения требований второго типа (система S_2). Так как неприоритетные требования не могут повлиять на обслуживание приоритетных, то S_2 является хорошо изученной системой $M/G/1/\infty$ с ненадежным прибором (см. раздел 2.1).

В системе S_1 мы имеем один прибор и два вида прерываний. Первый вид прерываний связан с поломкой прибора, а второй — с приходом приоритетного требования. Заметим, что эти прерывания могут возникнуть в любой момент.

Рассмотрим эти прерывания отдельно друг от друга. Прерывания i-го вида$(i = 1, 2)$ описываются последовательностями $\{u_k^i\}_{k=1}^{\infty}$ и $\{v_k^i\}_{k=1}^{\infty}$, где u_k^i— k-й интервал, в течение которого нет прерывания i-го вида, а v_k^i — время восстановления после k-го прерывания i-го вида. При $i = 1$ мы говорим о прерываниях, вызванных поломкой прибора, следовательно, u_k^1 распределено экспоненциально с параметром ν, а v_k^1 имеет распределение $D(x)$ и ПЛС $\delta(s)$. Если $i = 2$, то прерывание вызвано поступлением приоритетного требования, поэтому u_k^2 распределено экспоненциально с параметром λ_2, а v_k^2 — это k-й период занятости в системе S_2. Пусть $\pi(s)$ — ПЛС периода занятости в системе S_2 при условии, что период занятости начался в период рабочего состояния прибора. Из результатов [53](Kendall, 1951) следует, что функция $\pi(s)$ удовлетворяет уравнению

$$\pi(s) = \alpha_2(s + \lambda_2 - \lambda_2 \pi(s)), \tag{2.19}$$

где $\alpha_2(s)$ — ПЛС полного времени обслуживания в системе S_2 для требования, поступившего в занятую систему. В зависимости от дисциплины обслуживания требования после прерывания, мы можем найти $\alpha_2(s)$ по формуле (2.1) или (2.2).

2.4 Предельное распределение для числа требований в модели M_1

Система S_1 отличается от системы, рассмотренной в разделе 2.1, тем, что в S_1 есть два вида прерывания. Здесь мы покажем, как можно объединить эти два вида прерывания в один.

Будем говорить, что система S_1 находится в нерабочем состоянии, если прибор в данный момент сломан или есть хотя бы одно требование в системе S_2.

Таким образом, в S_1 возникает прерывание с интенсивностью ν из-за поломки прибора и с интенсивностью λ_2 из-за поступления приоритетного требования. Пусть $\{u_k\}_{k=1}^{\infty}$ — последовательность длин интервалов, в течение которых S_1 находится в рабочем состоянии, а $\{v_k\}_{k=1}^{\infty}$ — последовательность длин интервалов, в течение которых S_1 не работает. Из описания очевидно, что обе последовательности состоят из независимых одинаково распределенных случайных величин. При этом $u_1 \overset{d}{=} \min\{u_1^1, u_1^2\}$ ($\overset{d}{=}$ обозначает равенство по распределению), т.е. u_1 распределено экспоненциально с параметром $(\lambda_2 + \nu)$. Найдем ПЛС $\mathcal{X}(s)$ для случайной величины v_1.

Введем обозначения: $\tilde{u}_1^2 = u_1^2 - u_1^1, U_n^2 = \tilde{u}_1^2 + \sum_{k=2}^{n} u_k^2, V_n^2 = \sum_{k=1}^{n} v_k^2$. Заметим, что $\mathsf{P}(\tilde{u}_1^2 > x | u_1^1 < u_1^2) = e^{-\lambda_2 x}$.

Если прибор сломался раньше поступления приоритетного требования ($u_1^1 \leqslant u_1^2$) и за время ремонта прибора v_1^1 произошло n прерываний второго вида, т.е. пришло n приоритетных требований ($U_n^2 \leqslant v_1^1 < U_{n+1}^2$), то $v_1 = v_1^1 + V_n^2$. Наоборот, если поступление приоритетного требования произошло раньше, чем сломался прибор ($u_1^1 > u_1^2$), то $v_1 = v_1^2$. Справедлива формула

$$v_1 \overset{d}{=} \begin{cases} v_1^1 + V_n^2, & \text{если}((u_1^1 \leqslant u_1^2) \cap (U_n^2 \leqslant v_1^1 < U_{n+1}^2)),\ n = 0, 1, 2, \ldots; \\ v_1^2, & \text{если}(u_1^1 > u_1^2). \end{cases}$$

С её помощью найдем ПЛС $\mathcal{X}(s) = \mathsf{E}e^{-sv_1}$:

$$\mathcal{X}(s) = \mathsf{E}e^{-sv_1^2}1(u_1^1 > u_1^2) + \sum_{n=0}^{\infty} \mathsf{E}e^{-sv_1^1 - sV_n^2} 1(U_n^2 \leqslant v_1^1 < U_{n+1}^2) 1(u_1^1 \leqslant u_1^2).$$

Используя независимость случайных величин u_1^1, u_1^2, v_1^1, а также то, что период занятости в системе S_2 при работающем приборе имеет ПЛС $\pi(s)$, заключаем, что

$$\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2 + \nu}\pi(s) + \frac{\nu}{\lambda_2 + \nu}\delta(s + \lambda_2 - \lambda_2\pi(s)). \tag{2.20}$$

Так как u_1^1 и u_1^2 независимы и распределены экспоненциально, то несложно показать, что последовательности $\{u_n\}_{n=1}^{\infty}$ и $\{v_n\}_{n=1}^{\infty}$ независимы. В итоге мы свели исходную систему к системе из раздела 2.1. В нашем случае параметры принимают следующий вид:

$$\mu = \mu^* := \lambda_2 + \nu, \quad \zeta(s) = \zeta^*(s) := \mathcal{X}(s). \tag{2.21}$$

Пусть, как и раньше, q_n — количество требований в системе S_1 в момент ухода n-го требования. Тогда справедлив приводимый ниже результат, при доказательстве которого используется теорема (Бочаров, Печинкин, стр. 55) и сведение исходной системы к системе из раздела 2.1.

Следствие 5. *Если $\rho_1 = \lambda_1 a_1 < 1$, то существуют предельные (стационарные) вероятности состояний $p_i = \lim_{n\to\infty} \mathsf{P}(q_n = i), i = 0, 1, \ldots,$ и для функции $P_1(z) = \sum_{i=0}^{\infty} p_i z^i$ справедливо следующее соотношение:*

$$P_1(z) = \frac{\alpha_1(\lambda_1 - \lambda_1 z) - z\alpha_1^*(\lambda_1 - \lambda_1 z)}{\alpha_1(\lambda_1 - \lambda_1 z) - z} p_0, \tag{2.22}$$

где $p_0 = \frac{1-\rho_1}{1-\rho_1+\lambda_1 a_1^*}$; функции

$$\alpha_1(s) = \beta_1(s + \nu + \lambda_2 - (\nu + \lambda_2)\mathcal{X}(s)),$$
$$\alpha_1^*(s) = \frac{\lambda_1}{\lambda_1 + \lambda_2 - \lambda_2\pi(\lambda_1)} \left[1 + \lambda_2 \frac{\pi(s) - \pi(\lambda_1)}{\lambda_1 - s}\right] \alpha_1(s)$$

находятся с помощью формул (2.1), (2.3) и подстановки в последние формулы (2.21); a_1 и a_1^* — математические ожидания, соответствующие $\alpha_1(s)$ и $\alpha_1^*(s)$.

Если $\rho \geqslant 1$, то $q_n \xrightarrow{\mathsf{P}} \infty$ при $n \to \infty$.

Следствие 6. *Условие стабильности для системы S_1 имеет вид*

D1 $\quad (\lambda_1 b_1 + \lambda_2 b_2)(1 + \nu d) < 1,$

D2 $\quad \left(\frac{\lambda_1}{\lambda_2 + \nu}\frac{1 - \beta_1(\lambda_2 + \nu)}{\beta_1(\lambda_2 + \nu)} + \frac{\lambda_2}{\nu}\frac{1 - \beta_2(\nu)}{\beta_2(\nu)}\right)(1 + \nu d) < 1.$

Доказательство. Для получения условий стабильности в терминах исходных параметров нужно вычислить a_1. Сделаем это, используя формулы (2.4) и (2.5), получаем

$$\begin{aligned}
&\mathsf{D1} \qquad a_1 = b_1(1 + (\lambda_2 + \nu)(-\mathcal{X}'(0))),\\
&\mathsf{D2} \qquad a_1 = \frac{1 - \beta_1(\lambda_2 + \nu)}{\beta_1(\lambda_2 + \nu)} \left((-\mathcal{X}'(0)) + \frac{1}{\lambda_2 + \nu} \right).
\end{aligned}$$

В силу (2.20), имеем

$$-\mathcal{X}'(0) = \frac{\lambda_2}{\lambda_2 + \nu}\hat{\pi} + \frac{\nu}{\lambda_2 + \nu} d(1 + \lambda_2 \hat{\pi}),$$

где $\hat{\pi}$ — среднее время периода занятости в системе S_2. При помощи формулы (2.19), находим

$$\hat{\pi} = \frac{a_2}{1 - \rho_2},$$

где $\rho_2 = \lambda_2 a_2$ — коэффициент загрузки системы S_2. Далее найдем a_2 по формулам (2.4), (2.5)

$$\begin{aligned}
&\mathsf{D1} \qquad a_2 = b_2(1 + \nu d),\\
&\mathsf{D2} \qquad a_2 = \frac{1 - \beta_2(\nu)}{\beta_2(\nu)} \left(d + \frac{1}{\nu} \right).
\end{aligned}$$

Теперь мы можем в явном виде выписать значение a_1 и подставить в условие стабильности $\lambda_1 a_1 < 1$. ■

Если у времени обслуживания и ремонта есть первые k моментов, то предельное распределение процесса q_n имеет $k - 1$ моментов, которые можно вычислить по формуле (2.22).

Оценим для этой системы вероятность образования большой очереди с помощью результатов из раздела 2.2. Пусть $\Psi(x)$ — предельная функция распределения виртуального времени ожидания. Считаем, что при $x \to \infty$

$$1 - D(x) \sim \frac{(-1)^{m_1}}{\Gamma(1 - r_1)} x^{-r_1} K(x), \tag{2.23}$$

где $m_1 \in \mathbb{N}$ и $1 < m_1 < r_1 < m_1 + 1$, а $K(x)$ — медленно меняющаяся функция. Времена обслуживания также имеют "тяжелые хвосты"

$$1 - B_i(x) \sim \frac{(-1)^{l_i}}{\Gamma(1 - p_i)} x^{-p_i} L_1^{(i)}(x), \tag{2.24}$$

где $l_i \in \mathbb{N}$ и $1 < l_i < p_i < l_i + 1$, а $L_1^{(i)}$ — медленно меняющаяся функция, $i = 1, 2$.

Как было показано, при помощи подстановки (2.21) мы сводим нашу систему к системе из раздела 2.1. Пусть функция распределения $\hat{G}(x)$ имеет ПЛС $\mathcal{X}(s)$. Нужно найти асимптотику $1-\hat{G}(x)$ при $x \to \infty$ и затем применить теорему 2.

Лемма 5. *Если выполнены условия* (2.23) *и* (2.24), *то при* $x \to \infty$

$$1-\hat{G}(x) \sim \frac{(-1)^{n_1}}{\Gamma(1-q_1)}(1-\lambda_2 a_2)^{-q_1-1}x^{-q_1}K_2(x),$$

где $a_2 = b_1^2(1+\nu d), n_1 = \min(m_1, l_2), q_1 = \min(r_1, p_2)$, *а*

$$K_2(x) = 1(p_2 \leqslant r_1)\frac{\lambda_2}{\lambda_2+\nu}(1+\nu d)^{q_1+1}L_1^{(2)}(x) + 1(p_2 \geqslant r_1)\frac{\nu}{\lambda_2+\nu}K(x).$$

Доказательство. Воспользуемся формулой (2.20) и разложением в ряд для функций $\pi(s)$ и $\delta(s)$. Из работы [40](De Meyer, Teugels, 1980) следует, что при $s \to 0$

$$\pi_{n_1}(s) \sim s^{q_1}(1-\lambda_2 a_2)^{-q_1-1}K_1(1/s), \tag{2.25}$$

где $K_1(x) = 1(p_2 \leqslant r_1)(1+\nu d)^{q_1}L_1^{(2)}(x) + 1(p_2 \geqslant r_1)\nu b_2 K(x)$. Приведем разложение в ряд функции $\mathcal{X}(s)$

$$\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2+\nu}\left(\sum_{j=0}^{n_1}\frac{\hat{\pi}_j}{j!}(-s)^j + (-1)^{n_1+1}\pi_{n_1}(s)\right) +$$
$$+\frac{\nu}{\lambda_2+\nu}\left(\sum_{j=0}^{m_1}\frac{d_j}{j!}(-1)^j(s+\lambda_2-\lambda_2\pi(s))^j + (-1)^{m_1+1}\delta_{m_1}(s+\lambda_2-\lambda_2\pi(s))\right),$$

где $\hat{\pi}_j$, d_j — это j-ый момент соответствующий ПЛС $\pi(s)$ и $\delta(s)$. Так как $s+\lambda_2-\lambda_2\pi(s) \sim (1-\lambda_2 a_2)^{-1}s$ при $s \to 0$, то остаточный член разложения в ряд функции $\mathcal{X}(s)$ имеет вид

$$\mathcal{X}_{n_1}(s) = \frac{\lambda_2(1+\nu d)}{\lambda_2+\nu}\pi_{n_1}(s) + 1(p_2 \geqslant r_1)(1-\lambda_2 a_2)^{-r_1}\delta_{m_1}(s).$$

Отсюда, ввиду (2.25) и (2.23), следует утверждение леммы. ∎

Таким образом, положив

$$\begin{aligned}
B(x) &:= B_1(x), & \lambda &:= \lambda_1, & L_1(x) &:= L_1^{(1)}(x),\\
G(x) &:= \hat{G}(x), & \mu &:= \lambda_2+\nu, & L_2(x) &:= (1-\lambda_2 a_2)^{-q_1-1}K_2(x),\\
G^*(x) &:= \hat{G}(x), & \mu^* &:= \lambda_2+\nu, & L_2^*(x) &:= (1-\lambda_2 a_2)^{-q_1-1}K_2(x),
\end{aligned}$$

мы получаем ситуацию, описанную в разделе 2.2. Применив теорему 2, находим асимптотику $1-\Psi(x)$.

Следствие 7. *Если выполнены условия* (2.23) *и* (2.24), *то при* $x \to \infty$

$$1 - \Psi(x) \sim \frac{(-1)^{n_2-1}}{\Gamma(2-\hat{q}_2)} \frac{(1-\lambda_2 a_2)^{-q_2+1}}{1-(1+\nu d)(\lambda_1 b_1^1 + \lambda_2 b_1^2)} x^{-q_2+1} K_3(x),$$

где $n_2 = \min(l_1, n_1)$, $q_2 = \min(p_1, q_1)$ *и*

$$K_3(x) = 1(p_1 \leqslant q_1)\lambda_1(1+\nu d)^{p_1} L_1^{(1)}(x) + 1(p_1 \geqslant q_1)\frac{\lambda_2 + \nu}{1+\nu d} K_2(x).$$

Замечание 2. Если в нашем примере $\nu = 0$, то мы имеем систему обслуживания с надежным прибором и приоритетными требованиями. Такая система была рассмотрена в работе [27](Abate, 1997), где в случае "тяжелых хвостов"был получен результат, который совпадает с нашим, если в следствии 7 взять $\nu = 0$, а $r_1 = \infty$.

2.5 Описание модели M_2

Рассматривается одноканальная система массового обслуживания с ненадежным прибором и двумя типами поступающих требований. Прибор состоит из двух блоков, первый обслуживает требования первого типа, а второй — требования второго. При этом в любой момент времени может работать только один блок.

Входящие потоки — пуассоновские соответственно с параметрами λ_1 и λ_2. Времена обслуживания требований i-го типа ($i = 1, 2$) образуют последовательность $\{\eta_n^i\}_{n=1}^{\infty}$ независимых одинаково распределенных случайных величин с функцией распределения $B_i(x)$, средним b_i и ПЛС $\beta_i(s)$.

Каждый блок может выйти из строя во время обслуживания требования. Время рабочего состояния i-го блока ($i = 1, 2$) имеет экспоненциальное распределение с параметром μ_i, а время восстановления распределено по закону $G_i(x), \zeta_i(s) = \int_0^\infty e^{-sx} dG_i(x)$ и имеет математическое ожидание g_i.

Как и ранее, будем параллельно рассматривать две дисциплины поведения требования после прерывания D1 и D2.

Требования второго типа имеют абсолютный приоритет относительно требований первого типа. То есть, если во время обслуживания требования первого типа приходит требование второго типа, то обслуживание неприоритетного требования останавливается и начинается обслуживание приоритетного требования.

Модель можно рассматривать с точки зрения требований первого типа (эту систему будем обозначать S_1) и с точки зрения требований второго типа (система S_2). Так как неприоритетные требования не могут повлиять на обслуживание приоритетных, то S_2 является хорошо изученной системой $M/G/1/\infty$ с ненадежным прибором (см., раздел 2.1).

В системе S_1 мы имеем один прибор и два вида прерываний. Первый вид прерываний связан с поломкой первого блока, а второй вид — с приходом приоритетного требования. Заметим, что прерывание второго вида может возникнуть в любой момент.

Рассмотрим эти прерывания в предположении, что они независимы и никак не взаимодействуют. Тогда прерывания i-го вида$(i = 1, 2)$ описываются последовательностями $\{c_k^i\}_{k=1}^{\infty}$ и $\{d_k^i\}_{k=1}^{\infty}$, где c_k^i— k-ый интервал, в течении которого нет прерывания i-го вида, а d_k^i — время восстановления после k-го прерывания i-го вида. При $i = 1$ мы говорим о прерываниях вызванных поломкой первого блока, следовательно, c_k^1 распределено экспоненциально с параметром μ_1, а d_k^1 имеет распределение $G_1(x)$ и его ПЛС обозначим $\zeta_1(s)$. Если $i = 2$, то прерывание вызвано поступлением приоритетного требования, поэтому c_k^2 распределено экспоненциально с параметром λ_2, а d_k^2 это k-ый период занятости в системе S_2. Пусть $\pi(s)$ его ПЛС. Как мы упоминали ранее, функция $\pi(s)$ удовлетворяет уравнению

$$\pi(s) = \alpha_2(s + \lambda_2 - \lambda_2\pi(s)), \tag{2.26}$$

где $\alpha_2(s) = b_2(s + \mu_2 - \mu_2\zeta_2(s))$ — ПЛС полного времени обслуживания в системе S_2.

Из описания модели следует, что последовательности $\{c_k^i\}_{k=1}^{\infty}$, $\{d_k^i\}_{k=1}^{\infty}, i = 1, 2$ попарно независимы и каждая последовательность состоит из независимых одинаково распределенных случайных величин.

Мы рассмотрели три варианта того, как эти два вида прерывания взаимодействуют. Логично предположить, что во всех случаях, если произошло прерывание второго вида, то прерывание первого вида не возникает, то есть, если пришло приоритетное требование, то первый блок остается без работы, а, следовательно, сломаться не может. Теперь перечислим эти три варианта, которые обозначим символами V_i $(i = 1, 2, 3)$.

(V_1) Если во время ремонта первого блока приходит приоритетное требование, то ремонт приостанавливается на время периода занятости в S_2 и продолжается после ухода всех приоритетных требований.

(V_2) Аналогичен варианту V_1, с тем отличием, что ремонт не продолжается, а независимо начинается заново.

(V_3) Ремонт первого блока продолжается, если приходит приоритетное требование. Для преодоления технических трудностей в вычислениях дополнительно полагаем $G_1(x) = 1 - e^{-\nu x}$.

2.6 Предельное распределение для числа требований в модели M_2

Система S_1 отличается от системы, рассмотренной в разделе 2.1, тем, что в S_1 есть два вида прерывания. В данном разделе мы покажем, как можно объединить эти два вида прерывания в один и таким образом свести S_1 к хорошо изученной модели из раздела 2.1.

Будем говорить, что система S_1 находится в нерабочем состоянии, если первый блок в данный момент сломан или есть хотя бы одно требование в системе S_2.

Как и в упомянутой выше модели из [18](Бочаров, Печинкин, 1995), будем анализировать систему в двух случаях: когда S_1 свободна и занята.

Если S_1 свободна, то прерывание может возникнуть только из-за поступления приоритетного требования, а так как во время периода занятости в S_2 не может произойти поломка первого блока, то время восстановления имеет ПЛС $\pi(s)$. В терминах раздела 2.1, получаем, что в нашем случае

$$\mu^* = \lambda_2, \quad \zeta^*(s) = \pi(s). \tag{2.27}$$

Теперь предположим, что S_1 занята. Тогда прерывание может возникнуть с интенсивностью μ_1 от поломки первого блока и с интенсивностью λ_2 от поступления приоритетного требования. Пусть $\{c_k\}_{k=1}^{\infty}$ — последовательность длин интервалов, в течение которых S_1 находится в рабочем состоянии, а $\{d_k\}_{k=1}^{\infty}$ — последовательность длин интервалов, в течение которых S_1 не работает. Из описания очевидно, что обе последовательности состоят из независимых одинаково распределенных случайных величин. Причем $c_1 \stackrel{d}{=} \min\{c_1^1, c_1^2\}$, то есть c_1 распределено экспоненциально с параметром $(\lambda_2 + \mu_1)$ для всех вариантов $V_i, i = 1, 2, 3$. А вот распределение d_1 зависит от того, какой вариант взаимодействия мы рассматриваем. Найдем ПЛС $\mathcal{X}(s)$ случайной величины d_1 для каждого варианта взаимодействия.
Введем обозначения $\tilde{c}_1^2 = c_1^2 - c_1^1, C_n^2 = \tilde{c}_1^2 + \sum_{k=2}^{n} c_k^2, D_n^2 = \sum_{k=1}^{n} d_k^2$. Заметим, что $\tilde{c}_1^2$ имеет экспоненциальное распределение с параметром λ_2, если $c_1^1 < c_1^2$.

V_1 : Если первый блок сломался раньше прихода приоритетного требования $(c_1^1 \leqslant c_1^2)$ и за время ремонта первого блока d_1^1 произошло n прерываний второго вида $(C_n^2 \leqslant d_1^1 < C_{n+1}^2)$, то $d_1 = d_1^1 + \sum_{k=1}^{n} d_k^2$. Наоборот, если поступление приоритетного требования произошло раньше поломки первого блока $(c_1^1 > c_1^2)$, то $d_1 = d_1^2$. Для наглядности приведем рисунок 2.1 и укажем формулу, которой удовлетворяет d_1.

$$d_1 \stackrel{d}{=} \begin{cases} d_1^1 + \sum_{k=1}^{n} d_k^2 & ,\text{если}((c_1^1 \leqslant c_1^2) \cap (C_n^2 \leqslant d_1^1 < C_{n+1}^2)),\ n = 0, 1, 2, \ldots \\ d_1^2 & ,\text{если}(c_1^1 > c_1^2) \end{cases}.$$

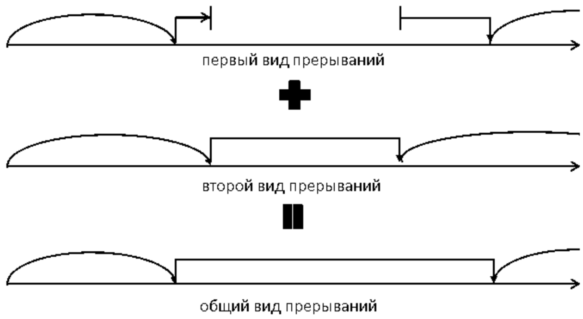

Рис. 2.1: общий вид прерываний

При помощи этого равенства найдем ПЛС $\mathcal{X}(s) = \mathsf{E}e^{-sd_1}$.

$$\mathcal{X}(s) = \mathsf{E}e^{-sd_1^2}1(c_1^1 > c_1^2) + \sum_{n=0}^{\infty} \mathsf{E}e^{-sd_1^1 - s\sum_{k=1}^n d_k^2}1(C_n^2 \leqslant d_1^1 < C_{n+1}^2)1(c_1^1 \leqslant c_1^2)$$

Из независимости c_1^1, c_1^2, d_1^1 и d_k^2 находим

$$\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2 + \mu_1}\pi(s) + \frac{\mu_1}{\lambda_2 + \mu_1}\zeta_1(s + \lambda_2 - \lambda_2\pi(s)).$$

V_2 : Пользуясь аналогичными рассуждениями, как и в варианте V_1, получаем формулы для d_1.

$$d_1 \stackrel{d}{=} \begin{cases} \tilde{d}_1 & ,\text{если}(c_1^1 \leqslant c_1^2) \\ d_1^2 & ,\text{если}(c_1^1 > c_1^2) \end{cases},$$

$$\tilde{d}_1 \stackrel{d}{=} \begin{cases} d_1^1 & ,\text{если}(d_1^1 < \tilde{c}_1^2) \\ \tilde{c}_1^2 + d_2^1 + \tilde{d}_1 & ,\text{если}(d_1^1 > \tilde{c}_1^2) \end{cases}. \tag{2.28}$$

Сначала находим ПЛС $\tilde{\mathcal{X}}(s)$ для случайной величины $\tilde{d}_1$ с помощью формулы (2.28).

$$\tilde{\mathcal{X}}(s) = \frac{\zeta_1(s + \lambda_2)}{1 - \pi(s)\frac{\lambda_2}{\lambda_2 + s}(1 - \zeta_1(s + \lambda_2))}.$$

После этого мы можем выписать формулу для ПЛС самой величины d_1.

$$\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2 + \mu_1}\pi(s) + \frac{\mu_1}{\lambda_2 + \mu_1}\frac{\zeta_1(s + \lambda_2)}{1 - \pi(s)\frac{\lambda_2}{\lambda_2 + s}(1 - \zeta_1(s + \lambda_2))}.$$

V_3 : Пользуясь теми же методами, что и в предыдущих вариантах, выпишем аналогичные формулы для варианта V_3.

$$d_1 \stackrel{d}{=} \begin{cases} \tilde{d}_1 & ,\text{если}(c_1^1 \leqslant c_1^2) \\ d_1^2 & ,\text{если}(c_1^1 > c_1^2) \end{cases},$$

$$\tilde{d}_1 \stackrel{d}{=} \begin{cases} d_1^1 & \text{, если}(d_1^1 \in \cup_{n=0}^{\infty}[C_n^2 + D_n^2, C_{n+1}^2 + D_n^2)) \\ C_n^2 + D_n^2 & \text{, если}(d_1^1 \in [C_n^2 + D_{n-1}^2, C_n^2 + D_n^2)), n = 1, 2, \ldots \end{cases}. \quad (2.29)$$

Для нахождения ПЛС d_1 нам достаточно найти ПЛС $\tilde{d}_1$. Опираясь на формулу (2.29), выпишем соотношение для $\tilde{\mathcal{X}}(s) = \mathsf{E}e^{-s\tilde{d}_1}$

$$\tilde{\mathcal{X}}(s) = \sum_{n=0}^{\infty} \mathsf{E}e^{-sd_1^1} 1(0 \leqslant d_1^1 - C_n^2 - D_n^2 < c_{n+1}^2) +$$
$$+ \sum_{n=1}^{\infty} \mathsf{E}e^{-s(C_n^2+D_n^2)} 1(0 \leqslant d_1^1 - C_n^2 - D_{n-1}^2 < d_n^2).$$

Воспользуемся предположением, что $G_1(x) = 1 - e^{-\nu x}$. Тогда первая сумма примет следующий вид

$$\sum_{n=0}^{\infty} \mathsf{E}e^{-sd_1^1} 1(C_n^2 + D_n^2 \leqslant d_1^1 < C_{n+1}^2 + D_n^2) =$$
$$= \sum_{n=0}^{\infty} \int_0^{\infty} \nu e^{-(s+\nu)x} (\mathsf{P}(C_n^2 + D_n^2 \leqslant x) - \mathsf{P}(C_{n+1}^2 + D_n^2 \leqslant x)) dx =$$
$$= \frac{\nu}{s+\nu} \sum_{n=0}^{\infty} \int_0^{\infty} e^{-(s+\nu)x} \left(d\,\mathsf{P}(C_n^2 + D_n^2 \leqslant x) - d\,\mathsf{P}(C_{n+1}^2 + D_n^2 \leqslant x) \right) =$$
$$= \frac{\nu}{s+\nu} \sum_{n=0}^{\infty} \left(\left(\frac{\lambda_2}{s+\lambda_2+\nu} \right)^n \pi^n(s+\nu) - \left(\frac{\lambda_2}{s+\lambda_2+\nu} \right)^{n+1} \pi^n(s+\nu) \right) =$$
$$= \frac{\nu}{s+\nu+\lambda_2(1-\pi(s+\nu))}.$$

А вторая сумма перепишется следующим образом

$$\sum_{n=1}^{\infty} \mathsf{E}e^{-s(C_n^2+D_n^2)} 1(C_n^2 + D_{n-1}^2 \leqslant d_1^1 < C_n^2 + D_n^2) =$$
$$= \sum_{n=1}^{\infty} \mathsf{E}e^{-s(C_n^2+D_n^2)} (e^{-\nu(C_n^2+D_{n-1}^2)} - e^{-\nu(C_n^2+D_n^2)}) =$$
$$= \frac{\lambda_2}{s+\lambda_2+\nu} (\pi(s) - \pi(s+\nu)) \sum_{n=2}^{\infty} \left(\frac{\lambda_2}{s+\lambda_2+\nu} \right)^n \pi^{n-1}(s+\nu)(\pi(s) - \pi(s+\nu)) =$$
$$= (\pi(s) - \pi(s+\nu)) \frac{\lambda_2}{s+\lambda_2+\nu} \left(\sum_{n=0}^{\infty} \left(\frac{\lambda_2}{s+\lambda_2+\nu} \pi^n(s+\nu) \right)^n \right) =$$
$$= \frac{\lambda_2(\pi(s) - \pi(s+\nu))}{s+\nu+\lambda_2(1-\pi(s+\nu))}.$$

В итоге $\tilde{\mathcal{X}}(s) = \frac{\nu+\lambda_2(\pi(s)-\pi(s+\nu))}{s+\nu+\lambda_2(1-\pi(s+\nu))}$, следовательно, получаем

$$\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2+\mu_1}\pi(s) + \frac{\mu_1}{\lambda_2+\mu_1}\frac{\nu+\lambda_2(\pi(s)-\pi(s+\nu))}{s+\nu+\lambda_2(1-\pi(s+\nu))}.$$

В итоге мы свели исходную систему к системе из раздела 2.1. В нашем случае параметры принимают следующий вид

$$\mu = \lambda_2+\mu_1, \quad \zeta(s) = \mathcal{X}(s). \tag{2.30}$$

Пусть q_n — количество требований в системе S_1 в момент ухода n-го требования. Тогда справедлива следующее утверждение.

Следствие 8. *Если $\rho_1 = \lambda_1 a_1 < 1$, тогда существуют предельные (стационарные) вероятности состояний $p_i = \lim\limits_{n\to\infty} \mathsf{P}\{q_n = i\}$ и для функции $P_1(z) = \sum_{i=0}^{\infty} p_i z^i$ справедливо следующее соотношение*

$$P_1(z) = \frac{\alpha_1(\lambda_1-\lambda_1 z) - z\alpha_1^*(\lambda_1-\lambda_1 z)}{\alpha_1(\lambda_1-\lambda_1 z) - z}p_0, \tag{2.31}$$

$$p_0 = \frac{1-\rho_1}{1-\rho_1+\lambda_1 a_1^*}, \tag{2.32}$$

где функции $\alpha_1(s)$ и $\alpha_1^*(s)$ находятся с помощью формул (2.1),(2.2) (2.3) *и подстановок туда* (2.27), (2.30). *Величины a_1 и a_1^* — математические ожидания, соответствующие $\alpha_1(s)$ и $\alpha_1^*(s)$.*

Если $\rho \geqslant 1$, то $q_n \xrightarrow[n\to\infty]{} \infty$ по вероятности.

Следствие 9. *Выпишем условия стабильности для каждого варианта $V_i(i = 1,2,3.)$ и дисциплин* D1, D2

$$\mathsf{D1}(V_1) \qquad \lambda_1 b_1(1+\mu_1 g_1) + \lambda_2 b_2(1+\mu_2 g_2) < 1,$$

$$\mathsf{D1}(V_2) \qquad \lambda_1 b_1\left(1+\mu_1\frac{1-\zeta_1(\lambda_2)}{\lambda_2\zeta_1(\lambda_2)}\right) + \lambda_2 b_2(1+\mu_2 g_2) < 1,$$

$$\mathsf{D1}(V_3) \qquad \lambda_1 b_1\left(1+\frac{\mu_1}{\nu+\lambda_2(1-\pi(\nu))}\right) + \lambda_2 b_2(1+\mu_2 g_2) < 1,$$

$$\mathsf{D2}(V_1) \qquad \frac{1-\beta_1(\lambda_2+\mu_1)}{\beta_1(\lambda_2+\mu_1)}\frac{\lambda_1}{\lambda_2+\mu_1}(1+\mu_1 g_1) + \frac{1-\beta_2(\mu_2)}{\beta_2(\mu_2)}\frac{\lambda_2}{\mu_2}(1+\mu_2 g_2) < 1,$$

$$\mathsf{D2}(V_2) \qquad \frac{1-\beta_1(\lambda_2+\mu_1)}{\beta_1(\lambda_2+\mu_1)}\frac{\lambda_1\left(1+\mu_1\frac{1-\zeta_1(\lambda_2)}{\lambda_2\zeta_1(\lambda_2)}\right)}{\lambda_2+\mu_1} + \frac{1-\beta_2(\mu_2)}{\beta_2(\mu_2)}\frac{\lambda_2}{\mu_2}(1+\mu_2 g_2) < 1,$$

$$\mathsf{D2}(V_3) \qquad \frac{1-\beta_1(\lambda_2+\mu_1)}{\beta_1(\lambda_2+\mu_1)}\frac{\lambda_1\left(1+\frac{\mu_1}{\nu+\lambda_2(1-\pi(\nu))}\right)}{\lambda_2+\mu_1} + \frac{1-\beta_2(\mu_2)}{\beta_2(\mu_2)}\frac{\lambda_2}{\mu_2}(1+\mu_2 g_2) < 1.$$

Доказательство. Все эти условия, по сути, являются подстановкой значения a_1 в условие $\lambda_1 a_1 < 1$ в каждом конкретном варианте. Алгоритм нахождения условий стабильности для двух дисциплин D1, D2 приведен в доказательстве следствия 6. ■

Пример 5. В качестве примера рассмотрим следующую модель.

Пусть у нас имеется завод,который получает заказы на изготовление двух типов приборов, λ_1 и λ_2 интенсивности поступления заказов первого и второго типа соответственно. Время изготовления прибора i-го $(i = 1, 2)$ типа имеет функцию распределения $B_i(x)$, ПЛС $\beta_i(s)$ и математическое ожидание b_i. Причем $\beta_2(s) = \frac{1}{1+b_2 s}$.

Для производства приборы второго типа имеют исключительную важность, поэтому, если завод работает над созданием прибора первого типа и поступает заказ на изготовление прибора второго типа, то работа по производству прибора первого типа приостанавливается (дисциплина D1) на время изготовления приборов второго типа (в нашей терминологии это вариант взаимодействия V_1). Причем для создания приборов первого типа используется старый станок, ломающийся с интенсивностью μ_1 и временем ремонта, распределенным по закону $G_1(x)$ с ПЛС $\zeta_1(s)$ и средним g_1. А для изготовления приборов второго типа используется новый станок, который не ломается $(\mu_2 = 0)$.

Пусть S_i $(i = 1, 2)$ - система обслуживания для изделий i-го типа.

Используя следствие 6 теоремы 8, выпишем условие эргодичности для процесса q_n системы S_1
$$\lambda_1 b_1(1 + \mu_1 g_1) + \lambda_2 b_2 < 1.$$

Для более глубокого анализа системы S_1, мы можем выписать явный вид производящей функции $P_1(z)$ предельного распределения q_n. Для этого необходимо найти $\alpha_1(s)$, $\alpha_1^*(s), a_1, a_1^*$ и подставить в формулы (2.31), (2.32). В итоге для этого нужно найти $\pi(s)$. Из-за того, что $\mu_2 = 0$, имеем
$$\alpha_2(s) = \frac{1}{1 + b_2 s}$$
и из (2.19), получаем
$$\pi^2(s) - \pi(s)\left(\frac{1 + b_2 s}{\lambda_2 b_2} + 1\right) + \frac{1}{\lambda_2 b_2} = 0.$$
Отсюда находим два корня, но из-за условия $\pi(0) = 1$, остается только один
$$\pi(s) = \frac{1 + b_2 s + \lambda_2 b_2 - \sqrt{(1 + b_2 s + \lambda_2 b_2)^2 - 4\lambda_2 b_2}}{2\lambda_2 b_2}.$$

При помощи этой формулы находим $\mathcal{X}(s) = \frac{\lambda_2}{\lambda_2+\mu_1}\pi(s) + \frac{\mu_1}{\lambda_2+\mu_1}\zeta_1(s + \lambda_2 - \lambda_2\pi(s))$. Далее, используя формулы (2.1),(2.3),(2.31),(2.32), мы находим явный вид функции $P_1(z)$.

Эта глава посвящена системам обслуживания с пуассоновским входящим потоком и ненадежным прибором. Рассмотрены обобщающие модели, в которых возможны две причины прерывания обслуживания: поступление приоритетного требования и поломка прибора. Для всех них найдены необходимые и достаточные условия эргодичности процессов виртуального времени ожидания и числа требований в системе. Для этих двух процессов получены явные выражения для предельных распределений в терминах преобразований Лапласа-Стилтьеса. Предположив, что времена ремонта и время обслуживания имеют распределения с регулярно меняющимися хвостами, найдена асимптотика вероятностей больших уклонений для процесса виртуального времени ожидания в стационарном режиме.

Глава 3

Вероятности больших уклонений для системы обслуживания с регенерирующим входящим потоком

Основной задачей, рассматриваемой в данной главе, является нахождение асимптотики вероятностей больших уклонений предельного (стационарного) распределения процесса виртуального времени ожидания. Мы решаем этот вопрос в рамках системы с регенерирующим входящим потоком и надежным прибором, в предположении, что функция распределения суммарной работы, поступившей в систему за период регенерации, имеет "тяжелый хвост"(результат для "легкого хвоста"приведен в [7](Афанасьева, Баштова, 2015)). Далее этот результат распространяется на случай, когда прибор ненадежен и время его ремонта имеет распределение с "тяжелым хвостом".

3.1 Описание модели

Рассматривается одноканальная система массового обслуживания с неограниченной очередью и регенерирующим входящим потоком $X(t)$ (определение см. на стр. 22), определенным на вероятностном пространстве $(\Omega, \mathscr{F}, \mathsf{P})$ с фильтрацией $\{\mathscr{F}_t, t \geqslant 0\}$, к которой он адаптирован. Процесс $X(t)$ имеет непрерывные справа неубывающие кусочно-постоянные траектории ($X(0) = 0$).

Случайные величины θ_j и $\tau_j = \theta_j - \theta_{j-1}$ ($j = 1, 2, \dots$) называются соответственно моментами и периодами регенерации. В теории очередей $X(t)$ представляет собой либо суммарное время обслуживания требований, либо число требований, поступивших в систему за время $(0, t]$. Здесь мы считаем, что $X(t)$ — суммарное время обслуживания поступивших требований, и обозначаем $\gamma_n = X(\theta_n) - X(\theta_{n-1})$ и ξ_n — число скачков $X(t)$ на n-ом периоде регенерации. Если в систему поступает ординарный поток требований, то ξ_n — это число пришедших требований, а величина каждого скачка — время обслуживания. Ввиду того что $X(t)$ регенерирующий поток, последовательности $\{\tau_j\}_{j=2}^{\infty}$, $\{\gamma_j\}_{j=2}^{\infty}$ и $\{\xi_j\}_{j=2}^{\infty}$ состоят из независимых одинаково распределенных случайных величин. Предположим, что $\mathsf{E}\gamma_2 < \infty$ и $\mathsf{E}\tau_2 < \infty$, и обозначим $G(y) = \mathsf{P}(\gamma_2 \leqslant y)$, $g(x) = \int_x^{\infty}(1 - G(y))dy$, $G^I(x) = \frac{1}{\mathsf{E}\gamma_2}\int\limits_0^x (1 - G(y))dy$.

Введем процесс виртуального времени ожидания $W(t)$, а также два вложенных процесса $W_n = W(\theta_n - 0)$ и $w_n = W(t_n - 0)$. Здесь t_n — момент n-го скачка процесса $X(t)$.

Для сходимости последовательности функций $\lim_{n\to\infty} \mathsf{P}(w_n \leqslant x) = F(x)$ при $n \to \infty$ достаточно потребовать выполнение условия апериодичности.

Условие 6. Наибольший общий делитель $\{k : \mathsf{P}(\xi_2 = k) > 0\} = 1$.

В работе [28](Афанасьева, Баштова, 2014) даны достаточные условия, при которых существуют пределы

$$\Phi(x) = \lim_{n\to\infty} \mathsf{P}(W_n \leqslant x),$$
$$\Psi(x) = \lim_{t\to\infty} \mathsf{P}(W(t) \leqslant x).$$

Более того, $\Psi(x)$, $\Phi(x)$ и $F(x)$ являются функциями распределения тогда и только тогда, когда *коэффициент загрузки*

$$\rho = \frac{\mathsf{E}\gamma_2}{\mathsf{E}\tau_2} < 1.$$

Далее везде предполагаем, что предельные распределения существуют и $\rho < 1$.

Наша основная задача — найти асимптотическое поведение функций $\Psi(x)$, $\Phi(x)$ и $F(x)$ при $x \to \infty$. Будем считать, что случайный вектор (τ, γ, ξ) имеет такое же распределение, как и вектор $(\tau_2, \gamma_2, \xi_2)$.

Следуя [44](Фосс и др. 2011), дадим определения классов функций распределения.

Определение 18. Функция распределения $F(x)$ называется асимптотически локально постоянной (принадлежит классу $\mathscr{L}$), если при $x \to \infty$

$$\bar{F}(x + y) \sim \bar{F}(x)$$

для любого фиксированного y. Для краткости такие функции будем называть просто локально постоянными.

Определение 19. Функция распределения $F(x)$ неотрицательной случайной величины принадлежит классу $\mathscr{S}$ субэкспоненциальных, если

$$\lim_{x\to\infty} \bar{F}^{*(2)}(x)/\bar{F}(x) = 2,$$

где $\bar{F}^{*(2)}(x) = \int_0^{\infty} \bar{F}(x - y)dF(y)$.

Функция распределения $F(x) = \mathsf{P}(\xi \leqslant x)$ на всей вещественной прямой называется субэкспоненциальной, если условное распределение $\mathsf{P}(\xi \leqslant x | \xi \geqslant 0)$ является субэкспоненциальным.

Известно [56](Kluppelberg, 1988), что любое субэкспоненциальное распределение является локально постоянным, т.е. $\mathscr{S} \subset \mathscr{L}$. В свою очередь, любое локально постоянное распределение имеет *тяжелый хвост*. Напомним, что случайная величина ξ имеет распределение с тяжелым хвостом, если $\mathsf{E}e^{\varepsilon\xi} = \infty$ для всех $\varepsilon > 0$.

Определение 20. Функция распределения $F(x)$ принадлежит классу $\mathscr{S}^*$ сильно субэкспоненциальных, если

$$\int_0^x \bar{F}(x-y)\bar{F}(y)dy \sim 2m\bar{F}(x),$$

где $m = \int_0^\infty \bar{F}(y)dy$.

Отметим, что одно из ключевых свойств сильно субэкспоненциального распределения является то, что само это распределение является субэкспоненциальным (т.е. $\mathscr{S}^* \subset \mathscr{S}$) и его интегральный хвост принадлежит классу $\mathscr{S}$ [56](Kluppelberg, 1988).

3.2 Основные теоремы

Далее, мы рассматриваем одноканальную систему обслуживания с регенерирующим входящим потоком $X(t)$.

Теорема 3. *Пусть $G^I(x) \in \mathscr{S}$ и при $y \to \infty$*

$$\mathsf{P}(\gamma > y + \tau) \sim \bar{G}(y). \tag{3.1}$$

Тогда при $x \to \infty$

$$\bar{\Phi}(x) \sim a^{-1}g(x), \tag{3.2}$$

где $a = \mathsf{E}\tau - \mathsf{E}\gamma$.

Доказательство. Из соотношения (3.1) следует, что интегральные хвосты $g(x) = \int\limits_x^\infty \bar{G}(y)dy$ и $\int\limits_x^\infty \mathsf{P}(\gamma > \tau + y)dy$ асимптотически эквивалентны (см., например, следствие 3.26 из [44](Фосс и др. 2011)).

Для доказательства (3.2) надо установить, что для любого $\varepsilon > 0$ найдется достаточно большое x_ε такое, что для всех $x > x_\varepsilon$

$$(1-\varepsilon)g(x)a^{-1} \leqslant \bar{\Phi}(x) \leqslant (1+\varepsilon)g(x)a^{-1}. \tag{3.3}$$

Оценка снизу. Введем вспомогательный процесс W_n^- с помощью рекуррентного соотношения

$$W_n^- = [W_{n-1}^- + \gamma_n - \tau_n]^+, \quad W_0^- = 0,$$

где $[x]^+ = \max(x, 0)$.

Очевидно, что если $W_0 = 0$ с вероятностью 1, то при всех $n = 1, 2, \ldots$ справедливо неравенство $W_n^- \leqslant W_n$. Обозначим $\bar{\Phi}^-(x) = \lim_{n\to\infty} \mathsf{P}(W_n^- > x)$. Из [71](Veraverbeke, 1977) известно, что если функция $\int_x^\infty \mathsf{P}(\gamma > y + \tau)dy$ субэкспоненциальна, то при $x \to \infty$ имеет место асимптотика

$$\bar{\Phi}^-(x) \sim a^{-1} \int_x^\infty \mathsf{P}(\gamma > y + \tau)dy \sim a^{-1} g(x).$$

Это даёт нижнюю оценку в (3.3).

Оценка сверху.

Лемма 6. *Пусть $G^I(x) \in \mathscr{S}$ и существует $T < \infty$ такое, что*

$$\mathsf{P}(\tau < T) = 1. \tag{3.4}$$

Тогда имеет место (3.2).

Доказательство. Как показано в [5](Афанасьева, Баштова, 2008), с вероятностью 1 выполняется неравенство

$$W_n \leqslant W_n^- + \tau_{k(n)},$$

где $k(n)$ — номер последнего до момента θ_n периода регенерации входного потока $X(t)$, на котором было освобождение системы. Используя (3.4), приходим к следующему неравенству

$$\bar{\Phi}(x) \leqslant \bar{\Phi}^-(x - T).$$

В свою очередь, $\bar{\Phi}^-(x-T) \sim a^{-1}g(x-T) \sim a^{-1}g(x)$ при $x \to \infty$, что вместе с оценкой снизу доказывает утверждение леммы. ■

Для доказательства теоремы нам понадобится вспомогательная система S_T, которая зависит от параметра T. Периоды регенерации $\{\tau_n^T\}_{n=1}^\infty$ входящего потока $X_T(t)$ определяются соотношением $\tau_n^T = \min(\tau_n, T)$, а моменты регенерации $\theta_n^T = \sum_{j=1}^n \tau_j^T$, $\theta_0^T = 0$. При этом, если $\tau_n < T$, то $X_T(t)$ на своем n-ом периоде регенерации совпадает с $X(t)$, то есть $X_T(\theta_{n-1}^T + t) = X(\theta_{n-1} + t)$ для $t \in (0, \tau_n]$. Если же $\tau_n > T$, то $X_T(t)$ до момента T копирует процесс $X(t)$, а все требования, что придут в основную систему за время $(\theta_{n-1} + T, \theta_n]$, поступят в S_T в момент $\theta_{n-1}^T + T$. Таким образом, для любого $t < T$ имеем

$X_T(\theta_{n-1}^T + t) = X(\theta_{n-1} + t)$, а при $t = T$ получаем $X_T(\theta_{n-1}^T + T) = X_T(\theta_n^T) = X(\theta_n)$. Выразим вышесказанное формулой

$$X_T(t) = \sum_{n=1}^{\infty} X(t - \theta_{n-1}^T + \theta_{n-1})\mathbb{I}(\theta_{n-1}^T < t \leqslant \theta_n^T),$$

где $\mathbb{I}(D)$ — индикатор события D.

Пусть $W_T(t)$ — виртуальное время ожидания для системы с входящим потоком $X_T(t)$, а $W_n^T = W_T(\theta_n^T - 0)$.

Лемма 7. *Если $W(0) \leqslant W_T(0)$ с вероятностью 1, то для любого $n \in \mathbb{N}$ с вероятностью 1 выполнено следующее неравенство*

$$W_n \leqslant W_n^T. \tag{3.5}$$

Доказательство. Имеем $W_0 \leqslant W_0^T$ и (3.5) выполнено для $n = 0$. Предположим, что (3.5) справедливо для $n-1$ и докажем для n. Рассмотрим n-ые периоды регенерации $(\theta_{n-1}, \theta_n]$ и $(\theta_{n-1}^T, \theta_n^T]$ для процессов $X(t)$ и $X_T(t)$ соответственно. Введем функцию $h(x, y) = \max(0, x + y)$, $x, y \in \mathbb{R}$.

Если $\tau_n \leqslant T$ или $\tau_n > T$, но в интервале $(\theta_{n-1} + T, \theta_n]$ процесс $X(t)$ не имеет скачков, то (3.5), очевидно, выполнено.

Предположим $\tau_n > T$ и в интервалах $(\theta_{n-1}, \theta_n]$ и $(\theta_{n-1}, \theta_{n-1} + T]$ процесс $X(t)$ имеет соответственно k и $l < k$ скачков. Обозначим $\{t_{n,j},\ j = \overline{1,k}\}$ — моменты скачков на n-ом периоде регенерации, $w_{n,j} = W(t_{n,j})$, $\eta_{n,j} = X(t_{n,j}) - X(t_{n,j} - 0)$ и $\zeta_{n,j} = t_{n,j} - t_{n,j-1}$, $(t_{n,0} = \theta_{n-1}), \zeta_{n,k+1} = \theta_n - t_{n,k}$.

Тогда имеем следующую систему равенств

$$\begin{aligned} w_{n,1} &= h(W_{n-1}, -\zeta_{n,1}), \\ w_{n,j} &= h(w_{n,j-1}, \eta_{n,j-1} - \zeta_{n,j}), \quad 2 \leqslant j \leqslant k, \\ W_n &= h(w_{n,k}, \eta_{n,k} - \zeta_{n,k+1}). \end{aligned} \tag{3.6}$$

Соответствующие обозначения для вспомогательной системы S_T пометим буквой T, то есть $\zeta_{n,j}^T = \zeta_{n,j}, j = \overline{1,l}, \zeta_{n,l+1}^T = T - t_{n,l}, \zeta_{n,j}^T = 0, j = \overline{l+2,k}$. Тогда соотношения (3.6) выполнены для $w_{n,j}^T$ и W_n^T.

Так как $h(x, y)$ возрастающая функция по x и y, $\zeta_{n,j}^T = \zeta_{n,j}$ при $j = \overline{1,l}$ и $W_{n-1} \leqslant W_{n-1}^T$, то очевидно

$$w_{n,j}^T \geqslant w_{n,j} \quad \text{при } j = \overline{1,l}.$$

Поэтому

$$\begin{aligned} W_n = h(w_{n,k}, \eta_{n,k} - \zeta_{n,k+1}) \leqslant w_{n,k} + \eta_{n,k} \leqslant \\ \leqslant w_{n,k-1} + \eta_{n,k-1} + \eta_{n,k} \leqslant \ldots \leqslant h(w_{n,l}, \eta_{n,l} - \zeta_{n,l+1}) + \eta_{n,l+1} + \cdots + \eta_{n,k} \leqslant \\ \leqslant h(w_{n,l}, \eta_{n,l} - \zeta_{n,l+1}^T) + \eta_{n,l+1} + \cdots + \eta_{n,k} \leqslant \\ \leqslant h(w_{n,l}^T, \eta_{n,l} - \zeta_{n,l+1}^T) + \eta_{n,l+1} + \cdots + \eta_{n,k} = W_n^T, \end{aligned}$$

что завершает доказательство леммы. ■

Найдем асимптотику при $x \to \infty$ предельного распределения $\lim_{n\to\infty} \mathsf{P}(W_n^T > x) = \bar{\Phi}_T(x)$, которое существует тогда и только тогда, когда $\rho_T = \frac{\mathsf{E}\gamma}{\mathsf{E}\tau^T} < 1$. Так как $\mathsf{E}\tau^T \xrightarrow[T\to\infty]{} \mathsf{E}\tau$, а $\frac{\mathsf{E}\gamma}{\mathsf{E}\tau} < 1$, то при достаточно большом T, мы имеем $\rho_T < 1$.

Чтобы применить лемму 6, нам достаточно показать, что функция распределения $G_T^I(y) = \frac{1}{\mathsf{E}\gamma^T}\int_0^x \mathsf{P}(\gamma^T > y)dy \in \mathscr{S}$ и $\mathsf{P}(\gamma^T > y + \tau^T) \sim \mathsf{P}(\gamma^T > y)$ при $y \to \infty$. Из конструкции вспомогательной системы имеем, что $\gamma_n^T = \gamma_n$ для любого n с вероятностью 1. Поэтому $G_T^I(y) = G^I(y) \in \mathscr{S}$. Второе условие вытекает из (3.1) и следующего неравенства для любого $y \geqslant 0$

$$\bar{G}(y) \geqslant \mathsf{P}(\gamma > y + \tau^T) \geqslant \mathsf{P}(\gamma > y + \tau).$$

Таким образом, из леммы 6 находим, что при $x \to \infty$

$$\bar{\Phi}_T(x) \sim a_T^{-1} g(x), \tag{3.7}$$

где $a_T = \mathsf{E}\min(\tau, T) - \mathsf{E}\gamma \to a$ при $T \to \infty$.

Для любого $\varepsilon > 0$ выберем T_ε из условия $a_{T_\varepsilon}^{-1} < a^{-1}(1+\varepsilon)$. В силу (3.7) найдется x_ε такое, что для всех $x > x_\varepsilon$

$$\bar{\Phi}_{T_\varepsilon}(x) \leqslant (1+\varepsilon) a_{T_\varepsilon}^{-1} g(x).$$

Учитывая (3.5), получаем

$$\bar{\Phi}(x) \leqslant \bar{\Phi}_{T_\varepsilon}(x) \leqslant (1+\varepsilon)^2 a^{-1} g(x).$$

Последнее неравенство вместе с нижней оценкой завершает доказательство теоремы. ■

Чтобы получить аналогичную асимптотику для функции $\bar{\Psi}(x)$, мы воспользуемся результатами работы [32](Asmussen, 1999). Если в условиях теоремы 3 дополнительно потребовать, чтобы функция распределения $\mathsf{P}(\gamma > y+\tau)$ была локально постоянной, то асимптотика (3.2) верна для $\bar{\Psi}(x)$.

Заметим, что если с вероятностью 1 процесс $X(t)$ имеет скачки в моменты регенерации, то асимптотика для $F(x)$ также следует из результатов [32](Asmussen, 1999). Достаточно рассмотреть случайное блуждание $S_n = X(t_n) - t_n = \sum\limits_{k=1}^{n}(\eta_k - \zeta_k)$, где t_n — момент n-го скачка, η_n — величина n-го скачка и $\zeta_k = t_{k+1} - t_k$. Тогда случайное блуждание S_n имеет регенерирующую структуру в смысле определения 1 и его точки регенерации $\{n_k\}_{k=1}^{\infty}$ таковы, что $t_{n_j} = \theta_j$ $(j = 1, 2, \dots)$. Однако, для многих регенерирующих процессов это условие не выполняется. Например, для полумарковского потока точки регенерации — это моменты попадания управляющего полумарковского процесса в некоторое фиксированное состояние и совсем не обязательно, чтобы в эти моменты поступали требования. Поэтому мы предложим другой подход, основанный на теореме 3. Введем функцию

$$v(x) = \frac{1}{\mathsf{E}\xi} \sum_{j=0}^{\infty} j\,\mathsf{P}(\gamma > x, \xi = j).$$

Теорема 4. *Пусть выполнены предположения теоремы 3, условие 6 и*

$$\lim_{t\to\infty} v(x)/g(x) = 0.$$

Тогда для $\bar F(x)$ имеет место (3.2).

Доказательство. Для времени ожидания n-го требования имеют место неравенства

$$\mathsf{P}(W_{r(n)-1} - \tau_{r(n)} > x) \leqslant \mathsf{P}(w_n > x) \leqslant \mathsf{P}(W_{r(n)-1} + \gamma_{r(n)} > x), \tag{3.8}$$

где $r(n)$ — номер периода регенерации, на котором пришло n-е требование.

Заметим, что $W_{r(n)-1}$ не зависит от $\tau_{r(n)}$ и $\gamma_{r(n)}$. Из усиленного закона больших чисел и эргодичности W_n несложно установить, что

$$\lim_{n\to\infty} \mathsf{P}(W_{r(n)} \leqslant x) = \Phi(x).$$

Ввиду апериодичности распределения ξ_2 (условие 6), из основной теоремы восстановления (см. на стр. 25) вытекает существование предела

$$\lim_{n\to\infty} \mathsf{P}(\xi_{r(n)} = j) = \frac{1}{\mathsf{E}\xi} j\,\mathsf{P}(\xi_1 = j). \tag{3.9}$$

Переходя в (3.8) к пределу при $n \to \infty$, находим при $x \geqslant 0$

$$\lim_{n\to\infty} \mathsf{P}(W_{r(n)-1} - \tau_{r(n)} > x) \leqslant \bar F(x) \leqslant \lim_{t\to\infty} \mathsf{P}(W_{r(n)-1} + \gamma_{r(n)} > x).$$

Оценим функцию $\bar F(x)$ сверху. Найдем предельное распределение $\gamma_{r(n)}$ при $n \to \infty$. Для этого выпишем

$$\mathsf{P}(\gamma_{r(n)} > x) = \sum_{j=0}^{\infty} \mathsf{P}(\gamma_{r(n)} > x | \xi_{r(n)} = j)\,\mathsf{P}(\xi_{r(n)} = j).$$

Так как по распределению $\mathsf{P}(\gamma_{r(n)} > x | \xi_{r(n)}) = \mathsf{P}(\gamma_2 > x | \xi_2)$ при любом n, из (3.9) получаем

$$\lim_{n\to\infty} \mathsf{P}(\gamma_{r(n)} > x) = \frac{1}{\mathsf{E}\xi} \sum_{j=0}^{\infty} j\,\mathsf{P}(\gamma_1 > x, \xi_1 = j) = v(x).$$

Далее нам понадобится следствие 2 из [65](Pitman, 1980). Если есть две независимые случайные величины δ_1 и δ_2, где δ_1 имеет субэкспоненциальное распределение и $\mathsf{P}(\delta_2 > x) = o(\mathsf{P}(\delta_1 > x))$ при $x \to \infty$, то при $x \to \infty$

$$\mathsf{P}(\delta_1 + \delta_2 > x) \sim \mathsf{P}(\delta_1 > x).$$

Заметим, что $\bar{\Phi}(x) \sim a^{-1}g(x)$ и $g(x)$ с точностью до нормировки — хвост субэкспоненциального распределения. По условию теоремы $v(x) = o(g(x))$. Поэтому при $x \to \infty$

$$\lim_{n\to\infty} \mathsf{P}(W_{r(n)-1} + \gamma_{r(n)} > x) \sim \bar{\Phi}(x) \sim a^{-1}g(x),$$

что обеспечивает оценку сверху.

Оценим функцию $\bar{F}(x)$ снизу. Зафиксируем произвольное $\varepsilon_2 > 0$ и выберем M_2 такое, что $\lim_{n\to\infty} \mathsf{P}(\tau_{r(n)} \leqslant M_2) > 1-\varepsilon_2$. Так как $\bar{\Phi}(x) \sim a^{-1}g(x)$ и $g(x)$ локально постоянная функция, то существует x_{ε_2} такое, что при всех $x > x_{\varepsilon_2}$

$$\frac{\bar{\Phi}(x+M_2)}{a^{-1}g(x)} \geqslant 1 - \varepsilon_2.$$

Следовательно, для достаточно больших x справедливы неравенства

$$\lim_{n\to\infty} \mathsf{P}(W_{r(n)-1} > x + \tau_{r(n)}) \geqslant \lim_{n\to\infty} \mathsf{P}(W_{r(n)-1} > x + M_2, \tau_{r(n)} \leqslant M_2) \geqslant$$
$$\geqslant \lim_{n\to\infty} \mathsf{P}(W_{r(n)-1} > x + M_2)\, \mathsf{P}(\tau_{r(n)} \leqslant M_2) \geqslant \frac{(1-\varepsilon_2)^2}{a} g(x).$$

Это вместе с оценкой сверху и произвольности выбора ε_2, завершает доказательство. ■

Приведем следствие, которое упростит проверку условий в случае, когда количество поступившей работы за период регенерации не зависит от длины этого периода.

Следствие 10. *Если $G(x) \in \mathscr{L}$, $G^I(x) \in \mathscr{S}$ и γ не зависит от τ, то* (3.2) *имеет место для функций $\bar{\Phi}(x)$ и $\bar{\Psi}(x)$.*

Доказательство. Так как $G(y) \in \mathscr{L}$ и $G^I(x) \in \mathscr{S}$, то остается найти предел $\lim_{y\to\infty} \frac{\mathsf{P}(\gamma > y+\tau)}{\bar{G}(y)}$.

Из положительности τ заключаем, что $\frac{\mathsf{P}(\gamma>y+\tau)}{\bar{G}(y)} \leqslant 1$ при любых y. Кроме того, при любом $0 < M < \infty$

$$\mathsf{P}(\gamma > y + \tau) \geqslant \mathsf{P}(\gamma > y + \tau, \tau < M) \geqslant \mathsf{P}(\gamma > y + M)\, \mathsf{P}(\tau < M).$$

Следовательно, для любого $\varepsilon > 0$ найдутся M_ε и y_ε такие, что $\mathsf{P}(\tau < M_\varepsilon) > 1 - \varepsilon$ и $\mathsf{P}(\gamma > y + M_\varepsilon)/\,\mathsf{P}(\gamma > y) > 1 - \varepsilon$ при всех $y > y_\varepsilon$, откуда вместе с верхней оценкой следует требуемое утверждение. ■

3.3 Системы с независимыми временами обслуживания

Здесь мы предположим, что процесс $A(t)$, задающий число требований, поступивших в систему за $[0,t)$, является регенерирующим в смысле определения 3. Времена обслуживания $\{\eta_n\}_{n=1}^{\infty}$ образуют последовательность независимых одинаково распределенных случайных величин, не зависящих от $A(t)$. Тогда $X(t) = \sum\limits_{j=1}^{A(t)} \eta_j$ также регенерирующий процесс с теми же моментами регенерации, что и $A(t)$. Интенсивность потока $A(t)$ определяется соотношением $\lambda = \lim_{t\to\infty} \frac{A(t)}{t} = \frac{\mathsf{E}\xi}{\mathsf{E}\tau}$, а коэффициент загрузки $\rho = \lambda b < 1$, где $b = \mathsf{E}\eta_j$. Пусть $B(x)$ — функция распределения времени обслуживания. Введем вспомогательную случайную величину $\tilde{\xi}$, которая имеет распределение $\mathsf{P}(\tilde{\xi} = k) = \frac{k}{\mathsf{E}\xi}\mathsf{P}(\xi = k)$, $(k = 1, 2, \dots)$.

Теорема 5. *Пусть $B(x) \in \mathscr{S}^*$, распределение ξ апериодично. Тогда верны следующие утверждения:*

(i) Если существует $c > b$ такое, что $\mathsf{P}(c\xi > x) = o(\bar{B}(x))$, то при $x \to \infty$

$$\bar{\Psi}(x) \sim \bar{\Phi}(x) \sim \frac{\lambda}{1-\lambda b}\int_x^{\infty} \bar{B}(y)dy. \tag{3.10}$$

(ii) Если существует $c > b$ такое, что $\mathsf{P}(c\tilde{\xi} > x) = o(\bar{B}(x))$, то асимптотика (3.10) *справедлива и для функции $\bar{F}(x)$.*

Доказательство.Пункт (i). Воспользуемся теоремой 1 из [39](Денисов и др. 2010), где утверждается, что если $B(y) \in \mathscr{S}^*$ и существует $c > b$ такое, что $\mathsf{P}(c\xi > x) = o(\bar{B}(x))$, то при $y \to \infty$

$$\bar{G}(y) = \mathsf{P}(\eta_1 + \dots + \eta_\xi > y) \sim \bar{B}(y)\mathsf{E}\xi. \tag{3.11}$$

Отсюда при $x \to \infty$

$$\frac{1}{a}g(x) \sim \frac{\mathsf{E}\xi}{\mathsf{E}\tau - \mathsf{E}\gamma}\int_x^{\infty} \bar{B}(y)dy = \frac{\lambda}{1-\lambda b}\int_x^{\infty} \bar{B}(y)dy.$$

Соотношение (3.11) означает, что $\bar{G}(y)$ и $\bar{B}(y)$ *пропорционально асимптотически эквивалентные* функции (см., определение 3.12 в [44](Фосс и др. 2011)). Поскольку $B(y) \in \mathscr{S}^*$, то $G(y)$ тоже сильно субэкспоненциальная функция распределения.

Далее проверим условие (3.1). Зафиксируем некоторые $M > 0$ и $N > 0$

$$\bar{G}(y) \geqslant \mathsf{P}(\gamma > y+\tau) \geqslant \mathsf{P}(\gamma > y+\tau, \tau < M) \geqslant \sum_{n=0}^{N} \mathsf{P}(\gamma > M+y, \xi = n, \tau < M) =$$

$$= \sum_{n=0}^{N} \mathsf{P}(\gamma > M + y|\xi = n)\,\mathsf{P}(\xi = n, \tau < M).$$

Для любого $\varepsilon > 0$ выберем M и N из условия

$$\sum_{n=1}^{N} n\,\mathsf{P}(\xi = n, \tau \leqslant M) > (1-\varepsilon)\mathsf{E}\xi.$$

По свойству субэкспоненциальных распределений, при любом конечном n, имеем

$$\mathsf{P}(\eta_1 + \cdots + \eta_n > y) \sim n\bar{B}(y) \quad \text{при } y \to \infty.$$

Поэтому можно подобрать $y_\varepsilon^{(1)}$ так, чтобы

$$\mathsf{P}(\eta_1 + \cdots + \eta_n > y) \geqslant n(1-\varepsilon)\bar{B}(y)$$

для любых $n \leqslant N$ и $y > y_\varepsilon^{(1)}$.

Так как $\bar{B}(y) \in \mathscr{L}$, то существует $y_\varepsilon^{(2)}$ такое, что

$$\bar{B}(y+M) \geqslant (1-\varepsilon)\bar{B}(y)$$

для любого $y > y_\varepsilon^{(2)}$. Пусть $y_\varepsilon = \max(y_\varepsilon^{(1)}, y_\varepsilon^{(2)})$. Тогда при $y > y_\varepsilon$ получаем

$$\mathsf{P}(\gamma > y + \tau) \geqslant \sum_{n=1}^{N} \mathsf{P}(\eta_1 + \cdots + \eta_n > M + y)\,\mathsf{P}(\xi = n, \tau < M) \geqslant$$
$$\geqslant (1-\varepsilon)\bar{B}(y+M)\sum_{n=1}^{N} n\,\mathsf{P}(\xi = n, \tau < M) \geqslant (1-\varepsilon)^3\bar{B}(y)\mathsf{E}\xi.$$

Следовательно, условие (3.1) выполнено.

Перейдем к пункту (ii). Так как из условия пункта (ii) следует условие (i), то (3.1) выполняется.

Чтобы применить теорему 4, нужно оценить при $x \to \infty$ поведение функции.

$$v(x) = \frac{1}{\mathsf{E}\xi}\sum_{k=1}^{\infty} k\,\mathsf{P}(\eta_1 + \cdots + \eta_\xi > x, \xi = k) = \mathsf{P}(\eta_1 + \cdots + \eta_{\tilde{\xi}} > x).$$

Используя теорему 1 из [39](Денисов и др. 2010), получаем, что при $x \to \infty$

$$v(x) \sim \frac{\mathsf{E}\tilde{\xi}}{\mathsf{E}\xi}\bar{B}(x).$$

Так как $g(x)$ субэкспоненциальная функция, то $v(x) = o(g(x))$, что и требовалось доказать. ■

3.4 Примеры

Пример 6 (Система с ненадежным прибором)**.** Рассмотрим систему с ненадежным прибором и регенерирующим входящим потоком. Прибор может выходить из строя только во время обслуживания. Время рабочего состояния распределено экспоненциально с параметром ν, а время восстановления прибора имеет функцию распределения $D(x)$, с преобразование Лапласа–Стилтьеса $d(s)$ и средним $d < \infty$. Если обслуживание требования было прервано, то после устранения прерывания требование обслуживается заново, причем новое время обслуживания не зависит от исходного (эта дисциплина D2 называется *preemptive-repeat-different*, см. раздел 2.1). Как уже отмечалось, в моделях с прерываниями весьма эффективно используется понятие *полного времени обслуживания*, введенное в [47](Gaver, 1962).Полное время обслуживания требования — это время, которое требование проводит на приборе вне зависимости от того, сломан прибор или нет. Обозначим $b(s) = \mathsf{E}e^{-s\eta_1}$ — преобразование Лапласа–Стилтьеса для времени обслуживания.

Пусть η_c — полное время обслуживания некоторого требования. В работе [47](Gaver, 1962) приведена формула для функции

$$b_c(s) = \mathsf{E}e^{-s\eta_c} = \frac{b(s+\nu)}{1 - d(s)\beta(s)(1-b(\nu))} \quad (Re\, s \geqslant 0), \tag{3.12}$$

где

$$\beta(s) = \frac{\nu}{1-b(\nu)} \int_0^\infty e^{-(s+\nu)y} \bar{B}(y)dy.$$

Из (3.12) находим

$$b_c = \mathsf{E}\eta_c = \frac{1-b(\nu)}{b(\nu)} \left(\frac{1}{\nu} + d\right) < \infty.$$

Условие эргодичности $\lambda b_c < 1$ считаем выполненным. Чтобы применить теорему 5, нам нужно найти асимптотику функции $\bar{B}_c(x) = \mathsf{P}(\eta_c > x)$ при $x \to \infty$. Формула (3.12) позволяет представить $b_c(s)$ в виде суммы

$$b_c(s) = \frac{b(s+\nu)}{b(\nu)} \sum_{j=0}^{\infty} (d(s)\beta(s))^j (1-b(\nu))^j b(\nu). \tag{3.13}$$

Из (3.13) следует представление

$$\eta_c \overset{d}{=} \tilde{\zeta} + \sum_{j=0}^{\mu} (v_j + \tilde{\eta}_j).$$

Здесь $\{\tilde\eta_j\}_{j=1}^\infty$, $\{v_j\}_{j=1}^\infty$ — независимые последовательности, состоящие из независимых одинаково распределенных случайных величин, а $\tilde\zeta$ и μ — независимые от этих последовательностей независимые случайные величины. При этом $\mathsf{P}(\mu=j)=b(\nu)(1-b(\nu))^j$, $j=0,1,\ldots,$ а

$$\mathsf{E}e^{-s\tilde\zeta}=\frac{b(s+\nu)}{b(\nu)},\ \mathsf{E}e^{-s\tilde\eta_1}=\beta(s),\ \mathsf{E}e^{-sv_1}=d(s).$$

Тогда

$$\mathsf{P}(\tilde\zeta>x)=\frac{1}{b(\nu)}\int_x^\infty e^{-\nu y}dB(y) \tag{3.14}$$

и

$$\mathsf{P}(\tilde\eta_1>x)=\frac{\nu}{1-b(\nu)}\int_x^\infty e^{-\nu y}\bar B(y)dy. \tag{3.15}$$

Полагая $D(x)\in\mathscr{S}^*$, оценим $\bar B_c(x)=\mathsf{P}(\eta_c>x)$ при $x\to\infty$. Используя формулу (3.15) и следствие 2 из [65](Pitman, 1980), находим при $x\to\infty$

$$\mathsf{P}(\tilde\eta_j+v_j>x)\sim\bar D(x).$$

Пусть $\tilde\mu$ — случайная величина с распределением $\mathsf{P}(\tilde\mu=j)=\mathsf{P}(\mu=j|\mu>0)=(1-b(\nu))^{j-1}b(\nu)$, $j=1,2,\ldots$. Из теоремы 1 в [39](Денисов и др. 2010) получаем при $x\to\infty$

$$\mathsf{P}\left(\sum_{j=1}^{\tilde\mu}(\tilde\eta_j+v_j)>x\right)\sim\mathsf{E}\tilde\mu\bar D(x)=\frac{1}{b(\nu)}\bar D(x).$$

Из формулы (3.14) видно, что $\tilde\zeta$ имеет легкий хвост, поэтому при $x\to\infty$

$$\bar B_c(x)=\mathsf{P}(\tilde\zeta>x,\mu=0)+\mathsf{P}(\tilde\zeta+\sum_{j=1}^{\mu}(\tilde\eta_j+v_j)|\mu>0)\,\mathsf{P}(\mu>0)\sim\frac{1-b(\nu)}{b(\nu)}\bar D(x).$$

Последнее соотношение показывает, что асимптотическое поведение полного времени обслуживания определяется функцией $\bar D(x)$, при условии сильной субэкспоненциальности $D(x)$. Теперь из теоремы 5 имеем

Следствие 11. *Пусть для системы $Reg/G/1$ с ненадежным прибором и дисциплиной $\mathsf{D2}$ функция распределения времени восстановления $D(x)\in\mathscr{S}^*$. Если $\mathsf{P}(r\xi>x)=o(\bar D(x))$ для некоторого $r>b_c$, то при $x\to\infty$*

$$\bar\Psi(x)\sim\bar\Phi(x)\sim\frac{\lambda(1-b(\nu))}{(1-\lambda b_c)b(\nu)}\int_x^\infty\bar D(y)dy.$$

Если $\mathsf{P}(r\tilde\xi>x)=o(\bar D(x))$, то эта асимптотика верна для $\bar F(x)$.

Замечание 3. В рассмотренной системе, асимптотика виртуального времени ожидания определяется функцией распределения времени восстановления прибора. Этот феномен обусловлен, в первую очередь, выбором дисциплины обслуживания требования после прерывания. Но если рассмотреть вариант, когда, например, после прерывания требование продолжает обслуживаться с момента остановки (такая дисциплина $\mathsf{D1}$ называется *preemptive-resume*), то асимптотика $\bar{B}_c(x)$ будет определяться функциями $\bar{D}(x)$ и $\bar{B}(x)$. В разделе 2.2 рассмотрена система $M/G/1$ с ненадежным прибором и дисциплиной $\mathsf{D1}$. Получена асимптотика для функции $\bar{B}_c(x)$ в предположении, что функции распределения $D(x)$ и $B(x)$ — регулярно меняющиеся (см., например, [44](Фосс и др. 2011))

$$\bar{B}_c(x) \sim \bar{B}\left(\frac{x}{1+\nu d}\right) + \nu b \bar{D}(x) \quad \text{при } x \to \infty.$$

Как видим, если хвост распределения $D(x)$ легче хвоста распределения $B(x)$, то функция распределения времени обслуживания будет определять асимптотическое поведение $\bar{\Phi}(x)$.

Пример 7 (Система с входящим ДСПП)**.** Рассмотрим одноканальную систему обслуживания с дважды стохастическим пуассоновским входящим потоком $A(t)$ (ДСПП), имеющим случайную интенсивность $\lambda(t,w)$ (см. подраздел 1.2.1), так что

$$A(t) = A^*(\Lambda(t)),$$

где $\Lambda(t) = \int_0^t \lambda(y,w)dy$ и $A^*(t)$ — стандартный пуассоновский поток, не зависящий от $\lambda(t,w)$. Определение и основные свойства ДСПП подробно описаны в книге [49](Grandell, 1976).

Времена обслуживания требований описываются последовательностью случайных величин $\{\eta_n\}_{n=1}^{\infty}$, которые независимы, имеют общую функцию распределения $B(x)$, среднее b, $\bar{B}(x) = 1 - B(x)$. Последовательность $\{\eta_n\}_{n=1}^{\infty}$ не зависит от процесса $A(t)$.

Пусть $\lambda(t,w)$ — регенерирующий процесс с точками регенерации $\{\theta_n\}(\theta_0 = 0)$, $\tau_n = \theta_n - \theta_{n-1}$ — длина n-го периода регенерации, $\mathsf{E}\tau < \infty$. Тогда $A(t)$ регенерирующий поток с теми же точками регенерации $\{\theta_n\}_{n=1}^{\infty}$, $\xi_n = A(\theta_n - 0) - A(\theta_{n-1})$ — количество требований, поступивших за n-ый период регенерации. Положим $\lim_{t\to\infty} \frac{A(t)}{t} = \frac{\mathsf{E}\xi}{\mathsf{E}\tau} = \lambda$. Выпишем функцию распределения случайной величины ξ

$$\mathsf{P}(\xi = n) = \mathsf{E}\frac{[\Lambda(\tau)]^n}{n!}e^{-\Lambda(\tau)}.$$

Полагаем, что коэффициент загрузки $\rho = \lambda b < 1$. Функции $\Phi(x)$, $\Psi(x)$ и $F(x)$ имеют тот же смысл, что и в предыдущих разделах.

Следствие 12. *Пусть* $B(x) \in \mathscr{S}^*$ *и*

1) $\lambda(t,w) < \tilde{\lambda} < \infty$ *с вероятностью 1;*

2) существует число $c > b$ *такое, что* $\mathsf{P}(\tau > x/c\tilde{\lambda}e^2) = o(\bar{B}(x))$ *при* $x \to \infty$.

Тогда для функций $\bar{\Psi}(x)$, $\bar{\Phi}(x)$, $\bar{F}(x)$ *имеет место* (3.10).

Доказательство. Как и ранее, введем случайную величину $\tilde{\xi}$ такую, что $\mathsf{P}(\tilde{\xi} = k) = \frac{k}{\mathsf{E}\xi}\mathsf{P}(\xi = k)$ для всех $k = \{1, 2, \dots\}$. В силу пункта (ii) теоремы 5, нам достаточно показать, что $\mathsf{P}(c\tilde{\xi} > x) = o(\bar{B}(x))$. Имеем

$$\mathsf{P}(\tilde{\xi} > x/c) \leqslant \int_0^\infty \sum_{k=[x/c]+1}^{\infty} \frac{(\tilde{\lambda}y)^k}{(k-1)!} e^{-\tilde{\lambda}y} d\,\mathsf{P}(\tau \leqslant y) =$$

$$\int_0^{x/(c\tilde{\lambda}e^2)} + \int_{x/(c\tilde{\lambda}e^2)}^\infty = J_1(x) + J_2(x).$$

Очевидно, что $J_2(x) \leqslant \mathsf{P}(\tau > x/(c\tilde{\lambda}e^2)) = o(\bar{B}(x))$ в силу условия 2).

Перейдем к оценке $J_1(x)$

$$J_1(x) = \int_0^{x/(c\tilde{\lambda}e^2)} \sum_{k=[x/c]+1}^{\infty} \frac{(\tilde{\lambda}y)^k}{(k-1)!} e^{-\tilde{\lambda}y} d\,\mathsf{P}(\tau \leqslant y) \leqslant$$

$$\leqslant \sum_{k=[x/c]+1}^{\infty} \frac{(\tilde{\lambda}m(x))^k}{(k-1)!} e^{-\tilde{\lambda}m(x)} = \tilde{\lambda}m(x)\,\mathsf{P}(\zeta \geqslant k(x)),$$

где $m(x) = \left[\frac{x}{c\tilde{\lambda}e^2}\right] + 1$, $k(x) = [x/c]$, а ζ имеет пуассоновское распределение с параметром $\tilde{\lambda}m(x)$. Пусть $\{\zeta_i\}_{i=1}^\infty$ последовательность независимых одинаково распределенных случайных величин, имеющих пуассоновское распределение с параметром $\tilde{\lambda}$. Тогда

$$\mathsf{P}(\zeta \geqslant k(x)) = \mathsf{P}(\zeta_1 + \dots + \zeta_{m(x)} \geqslant k(x)) = \mathsf{P}\left(\frac{\zeta_1 + \dots + \zeta_{m(x)}}{m(x)} \geqslant \frac{k(x)}{m(x)}\right).$$

Применим лемму 1.2 из [46](Ganesh et. al. 2004), в которой утверждается, что для последовательности независимых одинаково распределенных случайных величин $\{\zeta_i\}_{i=1}^\infty$ при любых x и n выполнено неравенство

$$\frac{1}{n}\ln \mathsf{P}\left(\frac{\zeta_1 + \dots + \zeta_n}{n} \geqslant x\right) \leqslant -\sup_{s \geqslant 0}(sx - \varphi(s)), \tag{3.16}$$

где $\varphi(s) = \ln \mathsf{E}e^{s\zeta_1}$. В нашем случае $\varphi(s) = \tilde{\lambda}(e^s - 1)$. Несложно установить, что

$$\sup_{s \geqslant 0}(sx - \varphi(s)) = x(\ln(x/\tilde{\lambda}) - 1) + \tilde{\lambda}.$$

Далее, из (3.16) получаем

$$\mathsf{P}\left(\frac{\zeta_1+\cdots+\zeta_n}{n}\geqslant x\right)\leqslant \exp(-n(x(\ln(x/\tilde\lambda)-1)-\tilde\lambda)).$$

Теперь применим этот результат к нашей ситуации. Подставив вместо $n := m(x)$ и $x := k(x)/m(x)$, находим

$$\mathsf{P}\left(\frac{\zeta_1+\cdots+\zeta_{m(x)}}{m(x)}\geqslant \frac{k(x)}{m(x)}\right)\leqslant \exp\left(-k(x)\left(\ln\left(\frac{k(x)}{m(x)\tilde\lambda}\right)-1\right)+\tilde\lambda m(x)\right).$$

Учитывая, что $m(x)=\left[\frac{x}{c\tilde\lambda e^2}\right]+1\leqslant(1+\delta)\frac{k(x)}{\tilde\lambda e^2}$ при любом $\delta>0$ и достаточно больших x, имеем

$$\exp\left(-k(x)\left(\ln\left(\frac{k(x)}{m(x)\tilde\lambda}\right)-1\right)+\tilde\lambda m(x)\right)\leqslant$$
$$\leqslant \exp\left(-k(x)\left(1-\ln(1+\delta)-\frac{1+\delta}{e^2}\right)\right).$$

Таким образом, вероятность $\mathsf{P}(\zeta>k(x))$ асимптотически убывает быстрее чем некоторая экспоненциальная функция, так что $J_1(x)=o(\bar B(x))$ и мы можем применить теорему 5. ■

Замечание 4. Если входящий поток является марковски модулированным и управляющий процесс имеет конечное число состояний, то можно не накладывать условия на интенсивность и на распределение длины периода регенерации, так как эти условия автоматически выполняются. Нужно лишь потребовать сильную субэкспоненциальность функции распределения времени обслуживания.

Пример 8 (Система с марковски модулированным полумарковским входящим потоком)**.** Здесь рассмотрим систему с марковски модулированным полумарковским входящим потоком (см., например, [29](Афанасьева и др. 2012)). Это ДСПП со случайной интенсивностью

$$\lambda(t,w)=\sum_{i=1}^{N}\lambda_i\mathbb{I}(U(t)=i),$$

где $\lambda_i\geqslant 0$, $i=1,\dots,N$, а $U(t)$ — полумарковский процесс. Он задается матрицей переходных вероятностей $P=\{P_{ij},\, i,j=\overline{1,N}\}$ и матрицей $\{C_{ij},\, i,j=\overline{1,N}\}$ функций распределений времен пребывания в состояний i при условии, что следующим состоянием будет $j\,(i,j=\overline{1,N})$. Если P — эргодическая матрица, то процесс $U(t)$ имеет предельное распределение

$\pi_i = \lim\limits_{t\to\infty} \mathsf{P}(U(t)=i)$, а его интенсивность $\lambda = \sum\limits_{i=1}^{N} \lambda_i \pi_i$. Времена обслуживания требований $\{\eta_n\}_{n=1}^{\infty}$ не зависят от входного потока и $B(x) = \mathsf{P}(\eta_n \leqslant x)$. В предположении, что $B(x) \in \mathscr{S}^*$ из следствия 12 несложно получить условие субэкспоненциальности функции $\Psi(x)$.

Положим

$$C(x) = \min_{1\leqslant i,j\leqslant N} C_{ij}(x).$$

Так как P — эргодическая матрица, то для некоторого $1 \leqslant m < \infty$ имеем $P_{ij}(m) > 0$ для всех $i, j = \overline{1, N}$. Здесь $\{P_{ij}(m)\} = P^m$. Не ограничивая общности, считаем $m = 1$. В качестве моментов регенерации входного потока рассмотрим моменты попадания управляющего полумарковского процесса $U(t)$ в некоторое фиксированное состояние, скажем, в первое. Тогда число переходов μ процесса $U(t)$ до возвращения в $\{1\}$ оценивается геометрически распределенной случайной величиной ξ с параметром $\alpha = \min_{i,j} P_{ij}$. Период регенерации τ стохастически удовлетворяет неравенству

$$\tau \leqslant \zeta_1 + \ldots \zeta_\mu,$$

где $\{\zeta_j\}_{j=1}^{\infty}$ — независимые одинаково распределенные случайные величины с функцией распределения $C(x)$, которые не зависят от μ.

Если $C(x)$ имеет легкий хвост, то (3.10) имеет место. Если $C(x)$ субэкспоненциальна, то в силу

$$\mathsf{P}(\tau > x) \leqslant \mathsf{P}(\zeta_1 + \cdots + \zeta_\mu > x) \sim \alpha^{-1}\bar{C}(x),$$

для (3.10) надо

$$\bar{C}(x/c\tilde{\lambda}e^2) = o(\bar{B}(x)) \tag{3.17}$$

для некоторого $c > b$. Таким образом, получено

Следствие 13. *Для системы с марковски модулированным полумарковским входящим потоком в предположениях* (3.17) *и* $B(x) \in \mathscr{S}^*$ *для функций* $\Psi(x)$, $\Phi(x)$ *и* $F(x)$ *имеет место* (3.10).

Замечание 5. Утверждение следствия 13 остается верным, если в моменты изменения состояния управляющего полумарковского процесса в систему поступают требования. Это почти очевидно, поскольку число таких требований на периоде регенерации оценивается сверху геометрически распределенной случайной величиной.

3.5 Заключение

Здесь мы обсудим влияние структуры входящего регенерирующего потока на асимптотическое поведение стационарного распределения $\Psi(x)$ виртуального времени ожидания. Рассмотрим систему с временами обслуживания, не зависящими от входного потока. В случае легкого хвоста распределения $B(x)$ в [7](Афанасьева, Баштова, 2015) установлено, что

$$\lim_{x\to\infty} \frac{1}{x} \ln \bar{\Psi}(x) = -q,$$

где q — единственное положительное решение уравнения

$$U(b(-q), q) = 1,$$

где $b(s) = \mathsf{E}e^{-s\eta}$, $U(z,s) = \mathsf{E}z^{\xi}e^{-s\tau}$, а η, τ, ξ соответственно время обслуживания, длина периода регенерации входного потока, число требований, поступивших за этот период. Мы видим, что асимптотическое поведение $\Psi(x)$ определяется совместным распределением ξ и τ и не зависит от того, как требования поступают в систему в течение периода регенерации. Ещё меньше асимптотика $\Psi(x)$ зависит от структуры входного потока в условиях теоремы 5, где она полностью определяется интенсивностью входного потока и интегральным хвостом функции распределения $B(x)$. Это происходит из-за предположения, что распределение ξ имеет более легкий хвост, чем распределение времени обслуживания. Если это не так, то, по-видимому, распределение ξ будет играть существенную роль. Продемонстрируем это следующим примером. Предположим, что длина периода регенерации имеет конечный носитель, то есть $\mathsf{P}(\tau \leqslant T) = 1$ для некоторого $T < \infty$, а функция распределения $C(x) = \mathsf{P}(\xi \leqslant x)$ регулярно меняющаяся. Как показано в [39](Денисов и др. 2010), если $\bar{B}(x) = O(\bar{C}(x))$, то при $x \to \infty$

$$\bar{G}(x) = \mathsf{P}(\eta_1 + \cdots + \eta_\xi > x) \sim \mathsf{E}\xi\bar{B}(x) + \bar{C}(x/b).$$

Поскольку интегральный хвост $G(x)$ субэкспоненциальный, то при $x \to \infty$

$$\bar{\Psi}(x) \sim \frac{\lambda}{1-\lambda b}\left[\int_x^\infty \bar{B}(y)dy + \frac{b}{\mathsf{E}\xi}\int_{x/b}^\infty \bar{C}(y)dy\right].$$

Заключение

Основные результаты диссертации состоят в следующем:

- Установлены необходимые и достаточные условия эргодичности процесса виртуального времени ожидания для одноканальной системы обслуживания с ненадежным прибором, в которой входящий поток требований и процесс, описывающий моменты поломок и ремонта прибора, являются регенерирующими и не зависят друг от друга.

- Для систем обслуживания $M/G/1/\infty$ с ненадежным прибором и наличием приоритетных требований получены стационарные распределения процессов виртуального времени ожидания и числа требований в терминах преобразования Лапласа-Стилтьеса. Найдена асимптотика вероятностей больших уклонений для процесса виртуального времени ожидания в стационарном режиме в случае, когда время обслуживания и времена ремонта имеют распределения с "тяжелыми хвостами". Результаты приведены для двух дисциплин обслуживания требования после восстановления прибора.

- Найдены достаточные условия субэкспоненциальности стационарного распределения виртуального времени ожидания в системе с регенерирующим входящим потоком и надежным прибором. Затем данный результат распространен на случай ненадежного прибора.

Дальнейшее исследование темы диссертации может быть связано с получением аналогичных результатов для многоканальной системы обслуживания с регенерирующим входящим потоком, где параметры функционирования каждого прибора определяются регенерирующими процессами. Большой интерес представляет анализ очереди, где прерывание обслуживания может происходить по многим причинам и источники прерываний взаимосвязаны.

Литература

[1] Л.Г. Афанасьева. О существовании предельного распределения в системах массового обслуживания с ограниченным временем пребывания. *Теория вероятностей и её применения*, 10(3):570–578, 1965.

[2] Л.Г. Афанасьева. О потоке потерянных требований в некоторых системах массового обслуживания с ограничением. *Изв. АН СССР Техн. кибернетика*, 3:57–65, 1966.

[3] Л.Г. Афанасьева, А.В. Мартынов. Об эргодических свойствах систем массового обслуживания с ограничением. *Теория вероятностей и её применения*, 14(1):105–114, 1969.

[4] Л.Г. Афанасьева. Об эргодичности открытой сети обслуживания. *Теория вероятностей и её применения*, 32(4):777–781, 1987.

[5] Л.Г. Афанасьева, Е.Е. Баштова. Предельные теоремы для систем обслуживания с дважды стохастическим пуассоновским потоком (условия высокой загрузки). *Проблемы передачи информации,* 44(4):352–369, 2008.

[6] Л.Г. Афанасьева. Системы массового обслуживания с циклическими управляющими процессами. *Кибернетика и системный анализ,* 41(1):54–68, 2005.

[7] Л.Г. Афанасьева, Е.Е. Баштова. Вероятности больших уклонений для системы обслуживания с регенерирующим входящим потоком. *Теория вероятностей и ее применения,* 60(1):171-177, 2015.

[8] Л.Г. Афанасьева, А.В. Ткаченко. Многоканальные системы обслуживания с регенерирующим входящим потоком. *Теория вероятностей и ее применения,* 58(2):210–234, 2013.

[9] Л.Г. Афанасьева, И.В. Руденко. Системы обслуживания $GI/G/\infty$ и их приложения к анализу транспортных моделей. *Теория вероятностей и ее применения,* 57(3):427—452, 2012.

[10] Л.Г. Афанасьева, Е.В. Булинская. Случайные процессы в теории массового обслуживания и управления запасами. *М:. Издательство МГУ,* 1980.

[11] А.А. Боровков. Некоторые предельные теоремы теории массового обслуживания. I. *Теория вероятн. и ее примен.,* 9(4):608–625, 1964.

[12] А.А. Боровков. О факторизационных тождествах и свойствах распределения супремума последовательных сумм. *Теория вероятн. и ее примен.,* 15(3):377–418, 1970.

[13] А.А. Боровков. Вероятностные процессы в теории массового обслуживания. *М.: Наука,* с. 367, 1972.

[14] А.А. Боровков. Теория вероятностей. *М.: Эдиториал УРСС,* с. 472, 1999.

[15] А.А. Боровков. Эргодичность и устойчивость случайных процессов. *Едиториал УРСС*, с. 440, 1999.

[16] А.А. Боровков. О субэкспоненциальных распределениях и асимптотике распределения максимума последовательных сумм. *Сибирский математический журнал,* 43(6):1253–1364, 2002.

[17] А.А. Боровков, К.А. Боровков. Вероятности больших уклонений для обобщенных процессов восстановления с правильно меняющимися распределениями скачков. *Математические труды,* 8(2):69-136, 2005.

[18] П.П. Бочаров, А.В. Печинкин. Теория массового обслуживания. *М. : Изд-во Рос. ун-та дружбы народов,* 1995.

[19] А.Н. Дудин, А.А. Назаров. Система обслуживания $MMAP/M/R/0$ с резервированием приборов, функционирующая в случайной среде. *Пробл. передачи информ.,* 51(3):93-104, 2015.

[20] Д.Г. Кендалл. Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова. *Математика,* 3(6):97–111, 1959.

[21] И.Н. Коваленко. Некоторые задачи массовго обслуживания с ограничением. *Теория вероятностей и её применения*, 6(2):222–228, 1961.

[22] Д. Кокс, В. Смит. Теория восстановления. *Советское радио*, с. 299, 1967.

[23] Т. Саати. Элементы теории массового обслуживания и её приложения. *М.: Советское радио,* 1971.

[24] Б.А. Севастьянов. Эргодическая теорема для марковских процессов и её приложение к телефонным линиям с отказами. *Теория вероятностей и её применения*, 2(1):106–116, 1957.

[25] В. Феллер. Введение в теорию вероятностей и её приложения, т. 2. *М.: Книжный Дом Либроком,* с. 752, 2010.

[26] А.Я. Хинчин. Работы по математической теории массового обслуживания. *Москва : Физматгиз,* Т. 236, 1963.

[27] J. Abate, W. Whitt. Asymptotics for M/G/1 low-priority waiting time tail probabilities. *Queueing Systems,* 25:173-233, 1997.

[28] L.G. Afanasyeva, E.E. Bashtova. Coupling method for asymptotic analysis of queues with regenerative input and unreliable server. *Queueing Systems,* 76(2):125–147, 2014.

[29] L.G. Afanasyeva, E.E. Bashtova, E.V. Bulinskaya. Limit theorems for semi-Markov queues and their applications. *Communications in Statistics-Simulation and Computation,* 41(6):688–709, 2012.

[30] S. Asmussen. Applied Probability and Queues. *Chichester: Wiley,* p.318, 1987.

[31] S. Asmussen. Ladder heights and the Markov-modulated $M|G|1$ queue. *Stochastic Processes and Their Applications*, 37:313–326, 1991.

[32] S. Asmussen. Semi-Markov queues with heavy tails. *Semi-Markov Models and Applications, J. Janssen and N. Limnios, Kluwer, Dordrecht*, 269–284, 1999.

[33] S. Asmussen, H. Schmidli, V. Schmidt. Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability,* 31(2):422–447, 1999.

[34] G. Alsmeyer, M. Sbignev. On the tail behaviour of the supremum of a random walk defined on a Markov chain. *Yokohama Math. J.*, 46:139–159, 1999.

[35] K. Arndt. Asymptotic properties of the distribution of the supremum of a random walk on a Markov chain. *Th. Prob. Appl.*, 25:309–324, 1980.

[36] N.H. Bingham, R.A. Doney. Asymptotic Properties of Supercritical Branching Processes I: The Galton-Watson Process. *Advances in Applied Probability*, 6(4):711-731, 1974.

[37] A.A. Borovkov, K.A. Borovkov. Asymptotic Analysis of Random Walks: Heavy-Tailed Distributions. *Cambridge University Press*, 2008.

[38] J.W. Cohen. The single server queue. *North Holland Publ. Co., Amsterdam,* p. 657, 1969.

[39] D. Denisov, S. Foss, D. Korshunov. Asymptotics of randomly stopped sums in the presence of heavy tails. *Bernoulli,* 16(4):971–994, 2010.

[40] A. De Meyer, J.L. Teugels. On the Asymptotic Behaviour of the Distributions of the Busy Period and Service Time in $M/G/1$. *Journal of Applied Probability,* 17(3):802-813, 1980.

[41] B.T. Doshi. Queueing systems with vacations — a survey. *Queueing systems,* 1(1):29–66, 1986.

[42] D. Fiems, T. Maertens, H. Brunee. Queueing systems with different types of interruptions. *Eur. J. Oper. Res.,* 188(3):838–845, 2008.

[43] S. Foss, T. Konstantopoulos, S. Zachary. Discrete and continuous time modulated random walks with heavy-tailed increments. *Journal of Theoretical Probability,* 20(3):581–612, 2007.

[44] S. Foss, D. Korshunov, S. Zachary. An introduction to heavy-tailed and subexponential distributions. *Springer, New York,* 2011.

[45] F.G. Foster. On the stochastic matrices associated with certain queuing processes. *Ann. Math. Statist.*, 24(2):355–360, 1953.

[46] A. Ganesh, N. O'Connel, D. Wischik. Big Queues. *Springer, Verlag Berlin Heidelberg,* 2004.

[47] D.P. Gaver. A waiting line with interrupted service including priority. *J Rl Stat Soc B*, 24:73–90, 1962.

[48] P.W. Glynn, W. Whitt. Logarithmic asymptotics for steadystate tail probabilities in a single-server queue. *Journal of Applied Probability, Special edition,* 31A:131—156, 1994.

[49] J. Grandell. Doubly stochastic Poisson processes. *Springer-Verlag,* 1976.

[50] R.A. Howard. Research in semi-Markovian decision structures. *J. Oper. Res. Soc Japan,* 6(4):163—-199, 1964.

[51] P. Jelenkovic, A. Lazar. Subexponential asymptotics of a Markovmodulated random walk with queueing applications. *J. Appl. Prob.,* 25:132—141, 1998.

[52] J. Keilson. Queues subject to service interruptions. *Ann Math Stat*, 33(4):1314–1322, 1962.

[53] D.G. Kendall. Some Problems in the Theory of Queues. *Journal of the Royal Statistical Society. Series B (Methodological),* 13(2):151–185, 1951.

[54] T. Kernane. A single server retrial queue with different types of server interruptions. *E-prints,* 2009.

[55] V. Klimenok, C.S. Kim, D. Orlovsky, A. Dudin. Lack of invariant property of Erlang loss model in case of the MAP input. *Queueing Systems*, 49:187–213, 2005.

[56] C. Kluppelberg. Subexponential distributions and integrated tails. *Journal of Applied Probability,* 25(1):132–141, 1988.

[57] A. Krishnamoorthy, P.K. Pramod, T.G. Deepak. On a queue with interruptions and repeat or resumption of service. *Nonlinear Anal. Theory Methods Appl.* 71(12):1673–1683, 2009.

[58] A. Krishnamoorthy, P. Pramod, S. Chakravarthy. Queues with interruptions: a survey. *TOP*, 1–31, 2012.

[59] W. Leland, M. Taqqu, W. Willinger, D. Wilson. On the self-similar nature of Ethernet traffic. *IEEE/ACM Trans. Netw.*, 2:1–15, 1994.

[60] D.V. Lindley The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press,* 48(2):277–289, 1952.

[61] R.M. Loynes. The stability of a queue with non-independent inter-arrival and service times. *Proc. Cambridge Philos. Soc.,* 58(3):497–520, 1962.

[62] Jr. R.G. Miller. Priority Queues. *Ann. Math. Statist.* 31(1):86–103, 1960.

[63] E. Morozov. The stability of a non-homogeneous queueing system with regenerative input. *J Math Sci.*, 83(3):407–421, 1997.

[64] A. Pakes. On the tails of waiting time distributions. *J. Appl. Prob.,* 7:745—789, 1975.

[65] E. Pitman. Subexponential distribution functions. *Journal of the Australian Mathematical Society (Series A),* 29(3):337–347, 1980.

[66] S.I. Resnick. Heavy tail modeling and teletraffic data. *Ann. Statist.,* 25:1805–1869, 1977.

[67] W.L. Smith. Regenerative stochastic processes. *Proc. Roy. Soc.*, A232:6–31, 1955.

[68] L.M. Takacs. Combinatorial methods in the theory of stochastic processes. *New York : Wiley,* v.126, 1967.

[69] H. Thorisson. The coupling of regenerative processes. *Adv. Appl. Probability,* 15:531–561, 1983.

[70] H. Thorisson. Coupling, stationarity, and regeneration. *New York : Springer,* v. 200, 2000.

[71] V. Veraverbeke. Asymptotic behaviour of Wiener-Hopf factors of a random walk. *Stochastic Processes and their Applications*, 5(1):27–37, 1977.

[72] H. White, L.S. Christie. Queuing with Preemptive Priorities or with Breakdown. *Operations Research,* 6(1):79–95, 1958.

[73] W. Willinger, M. Taqqu, W. Leland, D. Wilson. Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements. *Statistical Science,* 1995, 10:67–85, 1995.

Публикации автора

[74] С.Ж. Айбатов. Эргодическая теорема для системы обслуживания с ненадежным прибором. *Математические заметки*, 97(6):803–814, 2015.

[75] С.Ж. Айбатов. Система с приоритетным обслуживанием и ненадежным прибором. *Вестн. Моск. ун-та. Сер. 1 Матем. Механ.*, (3):39–42, 2016.

[76] С.Ж. Айбатов. Вероятности больших уклонений для систумы $M/G/1/\infty$ с ненадежным прибором. *Теория вероятн. и её примен.*, 61(2):378–384, 2016.

[77] S.Z. Aibatov. Queueing System with Preemptive Resume Service Discipline and Unreliable Server. *Journal of Mathematical Sciences*, 218(2):111–118, 2016.

[78] S.Z. Aibatov. Ergodic Theorem for a Single-Server Queue in a Random Environment. *XXXII International Seminar on Stability Problems for Stochastic Models,* Book of abstracts, June 16–21, Trondheim, Norway, 2014.

[79] С.Ж. Айбатов. Система с приоритетным обслуживанием и ненадежным прибором. *Всероссийский симпозиум по прикладной и промышленной математики, тезисы докладов,* 28 сентября – 5 октября, Сочи, 2014.

[80] С.Ж. Айбатов. Эргодическая теорема для одноканальных систем обслуживания с ненадежным прибором, функционирующих в случайной среде. *XXI Международная молодежная научная конференции студентов,*

аспирантов и молодых учёных «Ломоносов», Тезисы докладов, МАКС Пресс, Москва, 2014.

[81] S.Z. Aibatov. Queueing System with Preemptive Resume Service Discipline and Unreliable Server. *16th Applied Stochastic Models and Data Analysis International Conference (ASMDA2015),* Book of Abstracts, 8-9, ISAST University of Piraeus, Greece, 2015.

[82] С.Ж. Айбатов. Система с приоритетным обслуживанием и ненадежным прибором. *XXII Международная молодежная научная конференции студентов, аспирантов и молодых учёных «Ломоносов»,* Тезисы докладов, МАКС Пресс, Москва, 2015.

[83] S.Z. Aibatov. Waiting-time tail probabilities in queue with regenerative input flow and unreliable server. *VIII Moscow International Conference on Operations Research (ORM 2016)*, Conference Proceedings, Moscow, 1:202–205, 2016.

[84] С.Ж. Айбатов. Вероятности больших уклонений для системы обслуживания с входящим ДСПП и ненадежным прибором. *XXIII Международная молодежная научная конференции студентов, аспирантов и молодых учёных «Ломоносов»,* Тезисы докладов, МАКС Пресс, Москва, 2016.

[85] С.Ж. Айбатов, Л.Г. Афанасьева. Условия субэкспоненциальности стационарного времени ожидания в одноканальной системе обслуживания с регенерирующим входящим потоком. *Материалы международной конференции по стохастическим методам*, Тезисы докладов, с. 47, ЮФУ, Ростов-на-Дону, 2016.