

ГРИШЕЧКИН С. А.

ОЧЕРЕДИ И ВЕТВЯЩИЕСЯ ПРОЦЕССЫ*

1. Введение

Эта работа посвящена применениям теории ветвящихся случайных процессов в теории очередей. Основное свойство ветвящихся процессов — независимость жизней отдельных частиц (индивидуумов) — выделяет этот класс случайных процессов и позволяет строить обширную и хорошо проработанную теорию таких процессов. Достаточно полную информацию о проблемах, изучаемых этой теорией, можно получить из обзора [1] и монографий [12], [20], [34].

Мы будем использовать в этой работе несколько типов ветвящихся процессов. Наиболее общий из них — это ветвящиеся процессы Крампа–Мода–Ягерса, введенные в начале 70-х гг. О них можно прочитать в [34] и [35]. Частица x рождается в момент σ_x , живет время λ_x и в течение всей своей жизни может производить новые частицы. Этот процесс деторождения задается считающим процессом $\xi_x(\cdot)$: если $\xi_x(\cdot)$ имеет скачок величины n в точке t , то это значит, что частица x произвела n потомков в возрасте t .

В 1975 г. Ягерс ввел понятие *случайной характеристики*, которое будет широко использоваться в данной работе. Предполагается, что частица x в течение своей жизни производит некоторый продукт, и $\chi_x(t)$ — это количество продукта, произведенного частицей к моменту достижения ею возраста t (условимся, что $\chi_x(t) = 0$, если $t < 0$). Тогда $\chi_x(t - \sigma_x)$ — это продукт, произведенный частицей x к моменту t и $Z_t^\chi = \sum_x \chi_x(t - \sigma_x)$ — это общее количество продукта, произведенного всеми частицами к моменту t (в западной литературе Z_t^χ называется *branching process counted by random characteristic* χ). Независимость жизней отдельных частиц означает, что все тройки $(\lambda_x, \xi_x(\cdot), \chi_x(\cdot))$ есть независимые вероятностные копии $(\lambda, \xi(\cdot), \chi(\cdot))$. Мы будем, однако, иногда отходить от этого основополагающего требования, считая, что начальные (т.е. присутствующие в момент 0) частицы могут иметь другие законы распределения (на самом деле, это лишь незначительно

усложняет анализ).

Приведем несколько примеров. Будем обозначать $\mathbf{1}(\cdot)$ индикаторную функцию.

Пример 1.1. Пусть $\chi(t) = \mathbf{1}(t \in [0, \lambda))$. Тогда $Z_t^\chi = Z_t$ — число частиц, живущих в момент t .

Пример 1.2. Пусть $\chi(t) = t\mathbf{1}(t \in [0, \lambda)) + \lambda\mathbf{1}(t > \lambda)$. Тогда $Z_t^\chi = \int_0^t Z_u du$.

Пример 1.3. Пусть $\chi(t) = \mathbf{1}(t \geq 0)$. Тогда Z_t^χ — общее число частиц, рожденных до момента t .

Если каждая частица x производит все потомство в конце своей жизни (величина этого потомства есть $\xi_x = \xi_x(\infty)$), то такой процесс называется *процессом Севастьянова*. Если дополнительно потребовать независимость λ_x и ξ_x , то мы приходим к *процессу Беллмана–Харриса*. Процесс Беллмана–Харриса Z_t является марковским только в двух случаях: 1) λ_x есть константа или 2) λ_x имеет экспоненциальное распределение. В первом случае говорят о *процессе Гальтона–Ватсона*, во втором — о *марковском ветвящемся процессе с непрерывным временем*. Книга Б. А. Севастьянова [12] содержит наилучшее описание этих двух моделей.

Существуют различные модификации описанных моделей — они будут вводиться по мере необходимости.

По величине математического ожидания общего числа потомков одной частицы $\mathbf{E}\xi_x$ ветвящиеся процессы делятся на три класса: если $\mathbf{E}\xi_x < 1$, $\mathbf{E}\xi_x = 1$ или $\mathbf{E}\xi_x > 1$, то ветвящийся процесс называется, соответственно, *докритическим*, *критическим* или *надкритическим*. Докритический и критический процессы вырождаются почти наверное (т.е. найдется такой момент t , что $Z_t = 0$, если, разумеется, $\mathbf{P}\{\xi_x = 1\} < 1$); для надкритического случая типично экспоненциальное возрастание.

Первые применения ветвящихся процессов в теории очередей (см. [15], [21], [36]) были связаны с периодами занятости (промежутками времени между поступлением требования в пустую систему и первым моментом, когда система вновь освободится) в системе $\mathbf{M}/\mathbf{G}/1$, т.е. в однолинейной системе с пуассоновским входящим потоком и произвольным временем обслуживания одного требования. Требование, поступающее в пустую систему, интерпретируется как начальная частица, с которой начинается ветвящийся процесс. Требования, приходящие в систему во время обслуживания требования x , — это прямые потомки x . Их число ξ_x имеет производящую функцию $\mathbf{E}s^{\xi_x} = h(s) = \beta(\Lambda(1-s))$, где Λ — *интенсивность* входного потока требований, а β — преобразование Лапласа времени обслуживания одного требования. Все случайные величины ξ_x независимы и одинаково распределены, так как входящий поток является пуассоновским (то же верно для неординарного пуассоновского потока, поэтому излагаемые далее результаты могут быть обобщены и

на этот случай). Тогда общее число требований, обслуженных в течение периода занятости, совпадает с общим числом частиц, участвующих в ветвящемся процессе, т.е. с величиной $\lim_{t \rightarrow \infty} Z_t^\chi = Z_\infty^\chi$, где характеристика χ взята из [примера 1.3](#). Но каждый из ξ прямых потомков начальной частицы является основателем своего собственного ветвящегося процесса. Отсюда следует, что

$$Z_\infty^\chi \stackrel{d}{=} 1 + Z_\infty^\chi(1) + \dots + Z_\infty^\chi(\xi), \quad (1.1)$$

где символ $\stackrel{d}{=}$ обозначает совпадение распределений, $Z_\infty^\chi(i) \stackrel{d}{=} Z_\infty^\chi$ и случайные величины ξ , $Z_\infty^\chi(i)$, $i = 1, 2, \dots$, являются независимыми. Это соотношение может быть переписано в терминах производящих функций:

$$g(s) = sh(g(s)), \quad (1.2)$$

где $g(s) = \mathbf{E} s^{Z_\infty^\chi}$.

Аналогичным образом может быть найдено преобразование Лапласа продолжительности периода занятости. Для этого характеристику χ надо определить равенством $\chi_x(t) = l_x \mathbf{1}(t \geq 0)$, где l_x — время обслуживания требования x , $l_x \stackrel{d}{=} l$; $\mathbf{E} e^{-ul} = \beta(u)$, $l^{(1)} = \mathbf{E} l$. Аналогами (1.1) и (1.2) в этом случае будут уравнения

$$Z_\infty^\chi = l + Z_\infty^\chi(1) + \dots + Z_\infty^\chi(\xi), \quad (1.3)$$

$$g(u) = h(u, g(u)), \quad (1.4)$$

где $g(u) = \mathbf{E} \exp(-uZ_\infty^\chi)$ и $h(u, s) = \mathbf{E} e^{-ul} s^\xi = \beta(u + \Lambda(1 - s))$.

Отметим, что мы сознательно обошли вниманием вопрос, какая именно дисциплина обслуживания реализована в системе $\mathbf{M}/\mathbf{G}/1$ и с каким из видов ветвящихся процессов мы имели дело. С одной стороны, период занятости — один и тот же для любой консервативной дисциплины (у которой суммарная интенсивность обслуживания всех требований, присутствующих в системе, равна 1). С другой стороны, $Z_\infty^\chi = \sum_x \chi_x(\infty)$ и величина Z_∞^χ не зависит от того, как долго живет частица x (т.е. от λ_x), и от конкретного вида процесса деторождения $\xi_x(\cdot)$. Важны лишь общее количество потомков $\xi_x(\infty)$ и величина $\chi_x(\infty)$, т.е. ветвящиеся процессы с одним и тем же распределением $(\xi(\infty), \chi(\infty))$ будут иметь и совпадающие распределения Z_∞^χ .

Сделанное замечание должно внушать пессимизм: слишком мало величин в теории очередей, которые были бы инвариантны относительно дисциплины обслуживания, и почти нет результатов в теории ветвящихся процессов, которые не используют информацию о λ_x и $\xi_x(\cdot)$. Однако в дальнейшем будет показано, что связь между очередями и ветвящимися процессами гораздо теснее и не исчерпывается периодами занятости. Но сейчас мы приведем еще несколько результатов, не выходя за уже описанный круг вопросов.

Общее количество $W_{0,T}$ незавершенной работы в момент T (иными словами, виртуальное время ожидания при дисциплине FIFO — first in first out) также есть величина, не зависящая от конкретного вида консервативной дисциплины обслуживания. Покажем, как может быть найдено предельное распределение W_0 для распределений $W_{0,T}$ при $T \rightarrow \infty$.

Идея состоит в том, чтобы рассматривать $W_{0,T}$ при дисциплине LCFS — PR (last come first served — preemptive resume), когда каждое вновь поступающее требование прерывает обслуживание предыдущего требования и само занимает обслуживающий прибор. Эта идея была использована в [50] для системы $\mathbf{GI}/\mathbf{G}/1$, и там же была показана связь с ветвящимися процессами для системы $\mathbf{M}/\mathbf{G}/1$. Наш подход несколько отличается от использованного в этой работе.

В силу регенерирующей структуры системы (см. [19, р. 126])

$$\mathbf{E} e^{-uW_0} = \Lambda(1 - \rho)(\Lambda^{-1} + \mathbf{E}I), \quad (1.5)$$

где $I = \int_0^a e^{-uW_{0,T}} dT$, $\rho = \Lambda l^{(1)}$, a — продолжительность периода занятости, $\Lambda^{-1} + \mathbf{E}a = \Lambda^{-1}(1 - \rho)^{-1}$ — средняя продолжительность цикла регенерации. Предположим, что в момент 0 приходит требование (обозначим его тоже 0), с которого начинается период занятости a . Это требование обслуживается до момента $\tau(1)$, когда приходит новое требование 1, которое прерывает обслуживание требования 0. Требование 1 начинает свой собственный период занятости $a_1 \stackrel{d}{=} a$. Аналог величины $W_{0,T}$ для этого нового периода занятости обозначим $W_{1,T}$ (в $W_{1,T}$ не входит незавершенная работа требования 0). В момент $\tau(1) + a_1$ вновь начинается обслуживание требования 0. Следующее прерывание его обслуживания происходит в момент $\tau(2) + a_1$, когда приходит требование 2, с которым мы свяжем величины a_2 и $W_{2,T}$, и т. д. Ясно, что величины $I_i = \int_0^{a_i} e^{-uW_{i,T}} dT$ независимы и одинаково распределены, $I_i \stackrel{d}{=} I$. Нетрудно понять, что справедливо равенство

$$I = \int_0^l e^{-ut} dt + e^{-u(l-\tau(1))} I_1 + \dots + e^{-u(l-\tau(\xi))} I_\xi, \quad (1.6)$$

где l — время обслуживания требования 0, а ξ — число прерываний обслуживания требования 0. Очевидно, что $\tau(1), \tau(2), \dots$ — это скачки пуассоновского процесса, который не зависит от l и от I_i , $i = 1, 2, \dots$. Взяв математическое ожидание от обеих частей (1.6), приходим к равенству

$$\mathbf{E}I = J + \Lambda J(\mathbf{E}I), \quad (1.7)$$

где $J = \mathbf{E} \int_0^l e^{-ut} dt = (1 - \beta(u))/u$. Сопоставляя (1.5) и (1.7), получим формулу Хинчина–Поллачека

$$\mathbf{E} e^{-uW_0} = \frac{1 - \rho}{1 - \Lambda u^{-1}(1 - \beta(u))}. \quad (1.8)$$

В работах [39]–[42] идеология теории ветвящихся процессов (но не конкретные результаты из этой теории) была применена для исследования периодов занятости в других системах массового обслуживания. Обзорная статья [42] дает представление о полученных в этом направлении результатах.