Информационные процессы, Том 9, № 1, 2009, стр. 25–29

ОБ АСИМПТОТИКЕ ХВОСТА ПЕРИОДА ЗАНЯТОСТИ ПРИ КРИТИЧЕСКОЙ ЗАГРУЗКЕ

С.Ф.Яшков

Институт проблем передачи информации им. А.А.Харкевича, Российская академия наук
Большой Каретный переулок, 19, 127994 Москва ГСП–4, Россия.
E-mail: yashkov@iitp.ru, fax: (7–495)650–05–79

Поступила в редколлегию 1 апреля 2009

Аннотация—Изучается поведение хвоста распределения периода занятости в консервативной системе обслуживания $\mathrm{M/G/1/\infty}$ при критической загрузке $\rho = 1$. Функция распределения времени обслуживания принадлежит классу правильно меняющихся функций. Исследованы два случая: конечный и бесконечный второй момент распределения времени обслуживания. Первый из этих случаев рассматривался ранее в [7], где не предполагалось, что распределение времени обслуживания является правильно меняющейся функцией.

1. ВВЕДЕНИЕ И ПОДГОТОВИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В данной статье изучается поведение хвоста распределения периода занятости в консервативной системе обслуживания $\mathrm{M/G/1/\infty}$ при критической загрузке $\rho = 1$. В систему поступает однородный пуассоновский поток требований интенсивности λ, длительности обслуживания являются независимыми и одинаково распределенными (н.о.р.) случайными величинами B, имеющими функцию распределения $B(x) = \mathsf{P}(B \le x)$ с преобразованием Лапласа–Стилтьеса (ПЛС) $\beta(s) = \mathbb{E}[\mathrm{e}^{-sB}]$ и моментами $\beta_i,\ i = 1, 2, \ldots$. Функция распределения периода занятости $\Pi(x) = \mathsf{P}(\Pi \le x)$ имеет ПЛС $\pi(s) = \mathbb{E}[\mathrm{e}^{-s\Pi}]$. Моменты распределения случайной величины Π обозначаются $\pi_i,\ i = 1, 2, \ldots$, загрузка системы обслуживания есть $\rho = \lambda\beta_1$. Условие $\rho < 1$ является необходимым и достаточным условием существования стационарного режима системы. Хорошо известна следующая

Теорема 1.1 ([5, 7, 9, 12]). *Преобразование Лапласа–Стилтьеса распределения периода занятости в консервативной системе M/GI/1 удовлетворяет функциональному уравнению Кендалла–Такача*

$$\pi(s) = \beta(s + \lambda - \lambda\pi(s)). \tag{1.1}$$

Уравнение (1.1) *определяет единственную функцию* $\pi(s)$, *аналитическую в полуплоскости* $\mathrm{Re}\ s > 0$, *в которой* $|\pi(s)| \le 1$. *Функция* $\pi(s)$ *является вполне монотонной. If* $\rho \le 1$, *то* $\pi(0+) = \Pi(+\infty) = p^* = 1$, *в противном случае* $\pi(0+) < 1$, $\Pi(+\infty) = p^*$, *где* p^* *есть единственный положительный корень уравнения* $p^* = \beta(\lambda - \lambda p^*)$ *в* $(0, 1)$.

Напомним некоторые свойства решения уравнения (1.1) [7, §4.2]. При $\rho < 1$ уравнение (1.1) имеет единственное решение $0 < \pi(s) \le 1$, причем $\pi(0) = 1$ и $\pi(s) \to 1$ при $s \to 0$. Это означает, что период занятости конечен с вероятностью единица. Дифференцируя (1.1) по s (и переходя к пределу при $s \downarrow 0$ с использованием правила Лопиталя), нетрудно получить

$$\mathbb{E}[\Pi] = \pi_1 = \beta_1/(1 - \rho), \tag{1.2}$$

из которого вытекает, что $\mathbb{E}[\Pi] < \infty$.

При $\rho = 1$ уравнение (1.1) имеет единственное решение $\pi(s)$, и $\pi(s) \to 1$ при $s \to 0$. Однако теперь $\mathbb{E}[\Pi] = \infty$.

Когда $\rho > 1$, уравнение (1.1) при $s = 0$ имеет два решения $\pi_1(0) = 1$ и $\pi_2(0) < 1$, из которых мы должны выбрать минимальное решение. При $s > 0$ это же уравнение имеет единственное решение $\pi(s)$, причем $\pi(s) \to \pi_2(0)$ при $s \to 0$. Отсюда вытекает, что период занятости принимает значение $+\infty$ с вероятностью $1-\pi_2(0)$. Иными словами, $1-\Pi(x)$, т.е. хвост распределения периода занятости, есть вероятность события, что период занятости закончится после момента времени x, *если он вообще кончится.* При $\rho > 1$ существуют три возможности:

1. Период занятости закончится до момента времени x.
2. Период занятости закончится после момента времени x.
3. Период занятости никогда не закончится.

Хвост распределения периода занятости определяет вероятность события 2.

Напомним также, что преобразование Лапласа (по t) нестационарной вероятности отсутствия требований в системе M/G/1 в момент t (т.е. вероятности события, что число требований $L(t)$ в момент t равно 0) $\mathsf{P}_{00}(t) = \mathsf{P}(L(t) = 0 | L(0) = 0)$ тесно связано с теоремой 1.1, поскольку

$$p_{00}(s) = \int_{0-}^{\infty} \mathrm{e}^{-st}\, \mathsf{P}_{00}(t)\, dt = \frac{1}{s + \lambda - \lambda\pi(s)} \tag{1.3}$$

(см. [7, §4.3] and [16,17]).

Поведение хвоста распределения периода занятости в системе M/G/1 при $\rho < 1$ хорошо изучено (см., например, классические книги [1,2,6,7]). Однако при критической загрузке $\rho = 1$ о хвосте распределения периода занятости известно немногое. Большинство известных фактов можно найти в [7, пункт 3 §4.3] (см. дополнительно [6, §5.6]). Целью данной статьи является заполнение этого пробела для случая правильно меняющихся (regularly varying) [4] функций распределения времени обслуживания $B(x)$. Будут рассмотрены два случая: $\beta_2 < \infty$ и $\beta_2 = \infty$.

2. АСИМПТОТИКА ХВОСТА ПЕРИОДА ЗАНЯТОСТИ

2.1. Второй момент времени обслуживания конечен

Справедлива

Теорема 2.1. *Если $B(x) \in RV(-a)$, $a > 2$* [1]*, то*

$$\mathsf{P}(\Pi > x) \sim \beta_1 \sqrt{\frac{x\beta_1}{2\pi\beta_2}}, \quad x \to \infty. \tag{2.1}$$

Доказательство. Изучим поведение функции $\pi(s)$ при $s \downarrow 0$. Положим $\pi(s) = 1 - \varepsilon$. Используя разложение ПЛС $\beta(s)$ в ряд Тейлора в окрестности точки s при малых s

$$\beta(s) = 1 - \beta_1 s + \frac{\beta_2}{2!} s^2 + o(s^2), \quad s \downarrow 0$$

(ср. с теоремой 12.6 в [10]) и подставляя это равенство в функциональное уравнение (1.1), приходим к равенству

$$1 - \varepsilon = 1 - \beta_1(s + \lambda\varepsilon) + \frac{\beta_2}{2!}(s + \lambda\varepsilon)^2 + o((s + \lambda\varepsilon)^2). \tag{2.2}$$

[1] Это означает, что хвост распределения $1 - B(x) \sim Cx^{-a}\ell(x)$, где C — константа, а $\ell(x)$ есть медленно меняющаяся функция при $x \to \infty$, см., например, [4,16]. Кроме того, если $a > 2$, то $\beta_2 = \mathsf{P}[B^2] < \infty$.

Поскольку $\rho = \lambda\beta_1 = 1$, то (2.2) сводится к

$$\beta_1 s = \frac{\beta_2}{2!}(s+\lambda\varepsilon)^2 + o((s+\lambda\varepsilon)^2).$$

Отсюда вытекает, что $s + \lambda\varepsilon = \sqrt{\frac{2\beta_1 s}{\beta_2}}\,[1+o(1)]$, т.е.

$$1-\pi(s) = \beta_1\sqrt{\frac{2\beta_1 s}{\beta_2}}\,[1+o(1)], \quad s\downarrow 0. \tag{2.3}$$

Применим теперь к равенству (2.3) тауберову теорему Бингхема и Дони (см. теорему А в [3], отраженную также в [4, теорема 8.1.6], в [16, теорема в сноске 36]) и в приложении к этой статье. Эта тауберова теорема, впервые доказанная в контексте суперкритических ветвящихся процессов Гальтона–Ватсона, дает характеризацию асимптотического поведения хвоста функции распределения из класса $RV(-a)$ в терминах остаточного члена разложения в ряд Тейлора ПЛС этой функции распределения при малых значениях аргумента s. Принимая во внимание, что $\Gamma(-1/2) = -2\sqrt{\pi}$ ($\Gamma(\cdot)$ есть гамма–функция), а медленно меняющаяся функция $\ell(x)$ равна 1, из асимптотического равенства (2.3) вытекает утверждение (2.1). □

Следствие 2.1. *ПЛС* $p_{00}(s)$*, представленное формулой* (1.3)*, имеет асимптотическое разложение*

$$p_{00}(s) = \sqrt{\frac{\beta_2}{2\beta_1 s}} + o\left(\frac{1}{\sqrt{s}}\right), \quad s\downarrow 0. \tag{2.4}$$

Равенство (2.4) *эквивалентно*

$$\mathsf{P}(L(t)=0|L(0)=0) = \sqrt{\frac{\beta_2}{2\pi\beta_1 t}} + o\left(\frac{1}{\sqrt{t}}\right), \quad t\to+\infty. \tag{2.5}$$

Замечание 2.1. В книге [7] при доказательстве следствия 2.1 использовалась другая тауберова теорема из [8, с. 133], подобная теореме из статьи [11]. Эти теоремы связывают асимптотическое поведение функции распределения при $x\to\infty$ с ее ПЛС (аргумент s) в окрестностях полюсов ПЛС. В некотором смысле это более общие тауберовы теоремы, поскольку могут применяться для изучения поведения функций не только с полиномиальными хвостами, но и с экспоненциальными хвостами (или их комбинациями). Однако эти тауберовы теоремы нельзя применять в случаях, когда асимптотика описывается через медленно меняющиеся функции. В этом смысле использованная нами тауберова теорема является более мощной.

2.2. Второй момент времени обслуживания бесконечен

Справедлива

Теорема 2.2. *Если* $B(x)\in RV(-a),\ a\in(1,2)$ [2]*, то*

$$\mathsf{P}(\Pi > x) \sim \frac{1}{C\,\Gamma(1-1/a)}\left(\frac{\beta_1}{-\Gamma(1-a)}\right)^{1/a} x^{-1/a}, \quad x\to\infty, \tag{2.6}$$

где $\Gamma(\cdot)$ *есть гамма–функция.*

[2] Это означает, что хвост распределения $1-B(x)\sim Cx^{-a}\ell(x)$, где C — константа, а $\ell(x)$ есть медленно меняющаяся функция при $x\to\infty$ [4]. Кроме того, если $a\in(1,2)$, то $\beta_2 = \mathsf{P}[B^2] = \infty$.

Доказательство. В силу тауберовой теоремы Бингхема–Дони (см. приложение), ПЛС распределения $B(x)$ можно представить в виде

$$\beta(s) = 1 - \beta_1 s + (-\Gamma(1-a))Cs^a + o(s^a), \quad s \downarrow 0. \tag{2.7}$$

Подставляя равенство (2.7) в функциональное уравнение (1.1), можно получить асимптотическое равенство

$$1 - \pi(s) \sim \frac{1}{C}\sqrt[a]{\frac{\beta_1}{-\Gamma(1-a)}}\,\sqrt[a]{s}, \quad s \downarrow 0. \tag{2.8}$$

Повторное применение к равенству (2.8) тауберовой теоремы из приложения приводит к утверждению (2.6). □

Замечание 2.2. Теорема 2.2 позволяет сделать следующий интересный вывод: если $\rho = 1$, то чем тяжелее хвост распределения времени обслуживания, тем легче хвост распределения периода занятости.

Замечание 2.3. В силу консервативности системы обслуживания и равенства (1.3) доказанные утверждения дополняют известные из публикаций [13–15] результаты изучения стохастических систем с разделением процессора.

3. ЗАКЛЮЧЕНИЕ

Мы рассмотрели консервативную стохастическую систему M/G/1/∞, в которой функция распределения времени обслуживания принадлежит классу правильно меняющихся функций. Предполагалось, что система функционирует в условиях критической загрузки (отношение интенсивности входящего потока к интенсивности обслуживания равно единице). Исследовано асимптотическое поведение хвоста распределения периода занятости в такой системе обслуживания для случаев конечной и бесконечной дисперсии времени обслуживания. Основные результаты, представленные теоремами 2.1, 2.2 и замечанием 2.2, доказаны с помощью применения тауберовой теоремы Бингхема и Дони (1974).

ПРИЛОЖЕНИЕ

Тауберова теорема Бингхема и Дони [3, Th. A], [4, Th. 8.1.6, pp. 333–334].

Теорема приводится не в самой общей форме.

Теорема. *Пусть функция распределения $Q(x) = \mathsf{P}(Q \le x)$ имеет n моментов $q_1, \ldots, q_n$ и $q_0 = 1$. Предполагается, что $Q(x) \in RV(-a)$, $a > 0$ и имеет ПЛС $q(s)$. Определим*

$$d_n(s) = (-1)^{n+1}\left[q(s) - \sum_{i=0}^{n}(-s)^i/i!\right].$$

При $n < a < n+1$, $n = 1, 2, \ldots$ и константе $C > 0$, следующие утверждения эквивалентны:

$$d_n(s) = (C + o(1))s^a \ell(1/s), \quad s \downarrow 0,$$
$$1 - Q(x) = (C + o(1))(-1)^n x^{-a}\ell(x)/\Gamma(1-a), \quad x \to \infty,$$

где $\Gamma:\ (0,\infty) \to \mathbf{R}_+$ есть гамма-функция.

Случай целого a является более сложным. В сущности, приведенная теорема есть один из вариантов формального обоснования (для класса $RV(-a)$) первого принципа операционного исчисления, установленного Хевисайдом (Heaviside) в более общем случае.

СПИСОК ЛИТЕРАТУРЫ

1. Asmussen, S., *Applied Probability and Queues* (2nd edition), Heidelberg: Springer, 2003.
2. Bhat, U.N., *An Introduction to Queueing Theory. Modeling and Analysis in Applications*, New York: Springer, 2008.
3. Bingham, N.H. and Doney, R.A., Asymptotic Properties of Super–Critical Branching Processes. I, *Adv. Appl. Probab.*, 1974, vol. 6, no. 4, pp. 711–731.
4. Bingham, N.H., Goldie, C.M., and Teugels, J.L., *Regular Variation,* vol. 27 of *Encycl. of Math. and Its Appl.*, Cambridge: Cambridge Univ. Press, 1987.
5. Borel, E., Sur l'emploie du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au problème de l'attente à un guichet, *Comptes Rendus Hebd. des Séanc. de l'Académie des Sciences*, 1942, vol. 214, pp. 452–456.
6. Cox, D.R. and Smith, W.L., *Queues*, London: Methuen, 1961. Russian translation of 1966, Moscow.
7. Гнеденко Б.В., Коваленко И.Н. *Введение в теорию массового обслуживания* (1-ое изд.). М.: Наука, 1966. English translation of 1968 (1st edition), Jerusalem.
8. Диткин В.А., Прудников А.П. *Интегральные преобразования и операционное исчисление.* М.: Физматгиз, 1961.
9. Kendall, D.G., Some Problems in the Theory of Queues, *J. Roy. Statist. Soc., Ser. B*, 1951, vol. 13, pp. 151–185.
10. Kingman, J.F.C. and Taylor, S.J., *Introduction to Measure and Probability*, Cambridge: Cambridge Univ. Press, 1966.
11. Sutton, W.G.L., The asymptotic expansion of a function whose operational equivalent is known, *J. of the London Math. Soc.*, 1934, vol. 9, pp. 131–137.
12. Takács, L., *Introduction to the Theory of Queues*, Oxford: Oxford Univ. Press, 1962.
13. Yashkov, S.F., A Derivation of Response Time Distribution for an M/G/1 Processor-Sharing Queue, *Problems of Control and Info. Theory*, 1983, vol. 12, no. 2, pp. 133–148.
14. Yashkov, S.F., Processor-Sharing Queues: Some Progress in Analysis (Invited paper), *Queueing Syst.*, 1987, vol. 2, no. 1, pp. 1–17.
15. Яшков С.Ф. *Анализ очередей в ЭВМ.* М: Радио и связь, 1989.
16. Yashkov, S.F. and Yashkova, A.S., Processor Sharing: A Survey of the Mathematical Theory, *Autom. Remote Control*, 2007, vol. 68, no. 9, pp. 1662–1731.
17. Yashkov, S.F. and Yashkova, A.S., On asymptotics of the emptiness probability in the M/GI/1 queue, *Autom. Remote Control*, 2008, vol. 69, no. 3, pp. 525–528.

Статью представил к публикации член редколлегии В.И.Венец