ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ И МЕТОД КВАНТИЛЕЙ

Ю.И. Рыжиков (Санкт-Петербург)

Роль вероятностных процессов и методов их расчета в современной технике, да и во многих других областях жизни общества, невозможно переоценить. Одна из таких областей – теория очередей, она же – теория массового обслуживания (ТМО), имеющая разноообразные и многочисленные применения. Последующие рассуждения ориентированы главным образом на численные методы теории очередей, но могут быть использованы и в других областях – например, в теории связи.

Практически, все вероятностные расчеты основаны на выравнивании исходных и искомых распределений по методу моментов. Количество сохраняемых моментов рассматривается как порядок аппроксимации. По поводу метода моментов существует обширная литература [1]. Однако основной характеристикой работы СМО для ответственных приложений является вероятность завершения обслуживания за время, не превышающее директивного срока. Такая задача требует построения по ответным моментам соответствующей ДФР. Эта задача особенно актуальна для распределений с «толстыми хвостами» – медленно убывающей ДФР. Основной частью данного исследования являются технология и опыт работы с распределением Парето, имеющим *конечное* число моментов. Здесь возможности численных методов весьма ограничены, вследствие чего приходится прибегать к имитационному моделированию (ИМ). Его важнейшие достоинства – универсальность и наглядность, недостатки – большая трудоемкость, сложность отладки программ и принципиально ограниченная точность.

Ключевым элементом технологии ИМ является формирование псевдослучайных длительностей случайных величин на основе программных датчиков псевдослучайных чисел $\{U_i\}$, равномерно распределенных на интервале [0,1). Численные эксперименты (в частности, погрешности оценки π, определяемой через долю точек, попавших во вписанный в единичный квадрат круг) показали, что надежды на неограниченное увеличение точности при увеличении количества испытаний до нескольких миллионов явно не оправданы.

Самоподобные процессы и распределение Парето

Для трафика в информационно-вычислительных системах, задач управления дорожным движением и многих других приложений [1] весьма типичны распределения с «толстыми хвостами» (медленным убыванием плотности и ДФР) соответствующих распределений – настолько толстыми, что *высшие моменты этих распределений обращаются в бесконечность*. Особо отметим сохранение самоподобности при наложении источников трафика, т. е. при суммировании соответствующих потоков в мультиплексорах и коммутаторах. В [7] перечислены причины самоподобия телекоммуникационного трафика.

Из свойства самоподобия следует, что надежды на «сглаживание» трафика на длительных промежутках времени не имеют под собой никаких оснований. Вместо этого происходит не только разрежение и уплотнение потоков данных, но и кластеризация самих этих уплотнений. Всплески трафика происходят не поодиночке, а сериями. Считается, что общих аналитических результатов исследования очередей или влияния фрактальности трафика на качество его обслуживания в настоящее время не существует [5,6]. Однако по крайней мере часть интересующих практиков задач может быть решена с использованием *распределения Парето*. В частности, работа с ним весьма полезна для установления эффективности и границ применимости метода моментов.

Распределение Парето (оно же гиперболическое, или степенное [3-5,8]) имеет функцию распределения

$$F(t) = 1 - (K/t)^{\alpha}, \quad t \geq K. \tag{1}$$

Из этого выражения находим p -квантиль

$$t_p = K/(1-p)^{1\ \alpha}. \tag{2}$$

Момент распределения Парето порядка m

$$f_m = \int_K^{\infty} t^m \frac{\alpha K^{\alpha}}{t^{\alpha+1}} dt = \alpha K^{\alpha} \int_K^{\infty} t^{m-\alpha-1} dt = \alpha K^{\alpha} \cdot \left. \frac{t^{m-\alpha}}{m-\alpha} \right|_K^{\infty}$$

имеет конечное значение лишь при $\alpha > m$. В этом случае

$$f_m = \alpha K^{\alpha} \frac{K^{m-\alpha}}{\alpha - m} = \frac{\alpha K^m}{\alpha - m}. \tag{3}$$

При заданном α масштабный параметр K распределения Парето выражается через первый момент f_1 согласно

$$K = \frac{\alpha - 1}{\alpha} f_1. \tag{4}$$

Заметим, что из общих формул для прореживания и суммирования потоков (см.[2]) следует сохранение результирующими потоками свойства фрактальности.

Распределение Парето можно подбирать и *методом квантилей*. Из следует, что параметр α можно выразить через отношение дополнительных функций распределения:

$$\alpha = \frac{\mathrm{h}\,(\overline{F}_1/\overline{F}_2)}{\mathrm{h}\,(t_2/t_1)}. \tag{5}$$

Назначаемые квантили должны быть не слишком близкими друг к другу и допускать значительную вероятность их превышения.

Для выработки псевдослучайных чисел, подчиненных распределению Парето, воспользуемся методом обратной функции. Из условия

$$U = (K/T)^{\alpha}$$

получаем

$$T = K/U^{1\ \alpha}. \tag{6}$$

Здесь U – псевдослучайное число, равномерно распределенное между нулем и единицей. Генераторами распределения Парето снабжена популярная система ИМ GPSS/World. Расчет моментов сгенерированных последовательностей и вероятностей превышения заданных квантилей показывает, что оценки упомянутых вероятностей сходятся значительно быстрее, чем оценки моментов, причем независимо от параметра α – даже в общепризнанно опасной зоне $\alpha < 2$, где второго момента не существует.

Распределение Парето в расчете систем обслуживания

В статье [3] предложены по необходимости весьма громоздкие *численные* методы расчета одноканальных систем обслуживания систем Pa/M/1 и M/Pa/1. Не представляет труда обобщить методику [3] на случай GI/M/n.

С другой стороны, для системы M/G/1 из формулы Полячека – Хинчина следует, что для существования k-го момента ожидания необходима конечность $(k+1)$ -го момента. При Парето-обслуживании с параметром $\alpha < 2$ оценку даже первого момента ожидания (и пребывания в системе) невозможно получить ни численными, ни имитационными методами [3,8]. Для многоканальных систем с Парето-обслуживанием необходимо *специально организованное* имитационное моделирование.

Имитационное моделирование систем с Парето-обслуживанием

В табл. 1 сопоставляются результаты имитационного моделирования систем M/Pa/1 и M/Pa/3 при одинаковом коэффициенте загрузки (с утроением интенсивности входящего потока). Парето-обслуживание в одноканальных системах приводит к очень редким занятиям системы на чрезвычайно длительный срок. Можно предполагать, что случаи одновременного занятия «супердлинными» заявками всех каналов окажутся практически невероятными. Таким образом, при фиксированном коэффициенте загрузки дробление производительности вычислительной системы должно оказаться хорошей защитой от сверхдолгого ожидания. Здесь просматривается заметная тенденция к уменьшению как самих моментов распределения времени ожидания, так и их изменчивости по числу испытаний.

Таблица 1
Статистические моменты пребывания

Тыс. испыт.	M/Pa/1				M/Pa/3			
	$\alpha=1.5$		$\alpha=2.5$		$\alpha=1.5$		$\alpha=2.5$	
	v_1	v_2	v_1	v_2	v_1	v_2	v_1	v_2
2	1.482e2	3.425e4	5.260e0	6.357e1	3.247e1	1.923e3	2.011e0	6.491e0
5	6.371e1	1.380e4	4.484e0	4.201e1	1.407e1	7.739e2	1.834e0	5.237e0
10	5.120e2	8.531e5	5.630e0	9.716e1	5.674e1	1.067e4	1.996e0	6.693e0
20	5.929e2	8.383e5	4.880e0	6.694e1	5.115e1	8.862e3	1.952e0	6.723e0
50	2.772e2	3.463e5	4.615e0	5.759e1	2.814e1	3.965e3	1.902e0	6.493e0
100	1.534e2	1.766e5	4.208e0	4.431e1	1.619e1	2.033e3	1.916e0	7.220e0
200	9.421e1	9.393e4	4.314e0	4.906e1	1.504e1	1.846e3	1.902e0	6.703e0
500	5.785e1	4.280e4	4.295e0	4.672e1	9.822e0	8.852e2	1.877e0	6.337e0
1000	5.065e1	3.369e4	4.332e0	4.857e1	1.376e1	2.382e3	1.885e0	6.455e0
2000	1.669e2	6.995e6	4.327e0	4.862e1	1.110e2	1.446e5	2.002e0	2.026e1
5000	1.021e2	2.831e6	4.337e0	5.020e1	5.521e1	6.410e4	1.923e0	1.196e1
10000	9.393e1	1.491e6	4.484e0	7.606e1	4.287e1	4.077e4	1.914e0	9.903e0

В табл. 2 для тех же систем приводятся вероятности превышения заданных значений времени пребывания.

Таблица 2
Вероятности превышения квантиля q

Тыс. испыт.	M/Pa/1				M/Pa/3			
	$\alpha = 1.5$		$\alpha = 2.5$		$\alpha = 1.5$		$\alpha = 2.5$	
	$q = 3$	$q = 5$	$q = 3$	$q = 5$	$q = 3$	$q = 5$	$q = 3$	$q = 5$
2	8.11e-1	7.78e-1	5.07e-1	3.16e-1	7.50e-1	6.88e-1	2.02e-1	5.75e-2
5	6.41e-1	5.52e-1	4.93e-1	2.99e-1	4.01e-1	3.22e-1	1.54e-1	4.30e-2
10	7.51e-1	6.91e-1	5.05e-1	3.21e-1	5.82e-1	5.19e-1	1.86e-1	6.81e-2
20	8.14e-1	7.71e-1	4.78e-1	2.93e-1	6.13e-1	5.52e-1	1.70e-1	6.20e-2
50	7.02e-1	6.40e-1	4.61e-1	2.79e-1	5.06e-1	4.41e-1	1.56e-1	5.23e-2
100	6.08e-1	5.32e-1	4.43e-1	2.59e-1	4.07e-1	3.24e-1	1.50e-1	5.17e-2
200	5.69e-1	4.87e-1	4.45e-1	2.64e-1	3.74e-1	2.90e-1	1.55e-1	5.22e-2
500	5.59e-1	4.74e-1	4.46e-1	2.66e-1	3.43e-1	2.57e-1	1.51e-1	4.84e-2
1000	5.58e-1	4.71e-1	4.48e-1	2.69e-1	3.58e-1	2.73e-1	1.53e-1	4.81e-2
2000	5.62e-1	4.76e-1	4.46e-1	2.68e-1	4.01e-1	3.29-1	1.51e-1	4.77e-2
5000	5.58e-1	4.72e-1	4.47e-1	2.68e-1	3.66e-1	2.84e-1	1.50e-1	4.65e-2
10000	5.61e-1	4.75e-1	4.48e-1	2.69e-1	3.65e-1	2.83e-1	1.41e-1	4.66e-2

Стабилизация вероятностей превышения квантилей происходит значительно быстрее, чем стабилизация даже первого момента распределения времени пребывания, причем независимо от параметра α Парето-распределения обслуживания.

С помощью этих квантилей и вероятностей можно согласно и определить параметры Парето-распределения времени пребывания заявки в системе и далее (уже без дополнительной имитации) получить искомые вероятности для произвольных квантилей.

Заключение

Результатом данной работы является обоснованное предложение в имитационных экспериментах над моделями СМО с Парето-распределениями определять не моменты распределения времени пребывания заявки в системе, а вероятности превышения двух опорных квантилей. Это позволяет при разумном числе испытаний даже для $\alpha < 2$ построить ДФР упомянутого пребывания, дающую исследователю ответ на основной вопрос теории очередей. Отметим, что распределение времени пребывания в системе GI/M/n подчинено показательному закону независимо от типа входящего потока, что упрощает применение метода квантилей. Отмечена возможность исключения больших задержек при Парето-обслуживании путем дробления производительности системы.

Результаты могут быть использованы при анализе разнообразных ситуаций с нестандартной вариабельностью – прежде всего коммуникационных сетей.

Литература

1. **Ахиезер Н.И.** Классическая проблема моментов и некоторые вопросы анализа, связанные с нею. М.: Физматгиз, 1961. 310 с.
2. **Рыжиков Ю.И.** Алгоритмический подход к задачам массового обслуживания: Монография. СПб.: ВКА им. А. Ф. Можайского, 2013. 496 с.
3. **Рыжиков Ю.И.** Теория очередей и распределение Парето // Труды ВКА им. А.Ф. Можайского. 2015. Вып. 648. С. 28–43.
4. **Вадзинский Р.Н.** Справочник по вероятностным распределениям. СПб.: Наука, 2001. 295 с.
5. **Задорожный В.Н.** Аналитико-имитационные исследования систем и сетей массового обслуживания: Монография. Омск: изд-во ОмГТУ, 2010. 280 с.
6. **Задорожный В.Н.** Аналитико-имитационные исследования Больших Сетевых Структур: Монография. Омск: изд-во ОмГТУ, 2011. 208 с.
7. **Шелухин О.И.** Причины самоподобия телетрафика и методы оценки показателя Херста // Электротехнические и информационные комплексы и системы. 2007. № 1. Т.3.С.5–14.
8. **Кутузов О.И., Татарникова Т.М.** К оцениванию и сопоставлению очередей классических и фрактальных систем массового обслуживания // Информационно-управляющие системы. 2016. № 2. С. 48–55.