ЧИСЛЕННЫЕ МЕТОДЫ ТЕОРИИ ОЧЕРЕДЕЙ

Ю. И. Рыжиков (Санкт-Петербург)

Современная теория очередей по необходимости исследует сложные математические модели при весьма общих допущениях, что определяет и сложность полученных результатов. В связи с этим особую ценность приобретают *численные* методы теории очередей. Их теоретическими основами являются:

- аппроксимация исходных распределений интервалов между заявками и длительностей обслуживания более удобными для расчетов;
- «законы сохранения» (см. доклад того же автора о роли имитации в компьютерном моделировании);
- потокоэквивалентная декомпозиция сетей обслуживания.

Аппроксимация распределений

В качестве аппроксимирующих применяются:

- фазовые распределения с показательно распределенными задержками в фазе (эрлангово, гиперэкспоненциальное, коксово) – для расчета распределений вероятностей состояний сложных систем;
- гамма-распределение – для вычисления необходимых во многих задачах распределений числа заявок пуассоновского потока за случайный интервал времени;
- распределение Вейбулла – для построения функции распределения времени пребывания заявки в системе (это основной показатель, интересующий заказчика оперативных систем с повышенной ролью фактора времени) по ее моментам. Кроме того, распределение Вейбулла полезно при исследовании многоканальных систем (в этом случае ДФР времени освобождения хотя бы одного канала будет иметь распределение того же типа с пересчитанным масштабным параметром).

Приведем пример диаграммы одного из фазовых распределений:

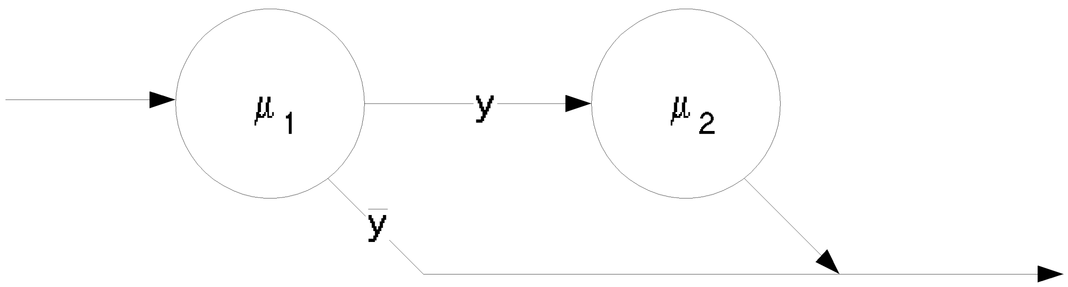

Рис. 1. Двухфазное распределение Кокса

Соответственно можно нарисовать диаграмму переходов между микросостояниями системы по завершению обслуживания, длительность которого подчинена распределению Кокса:

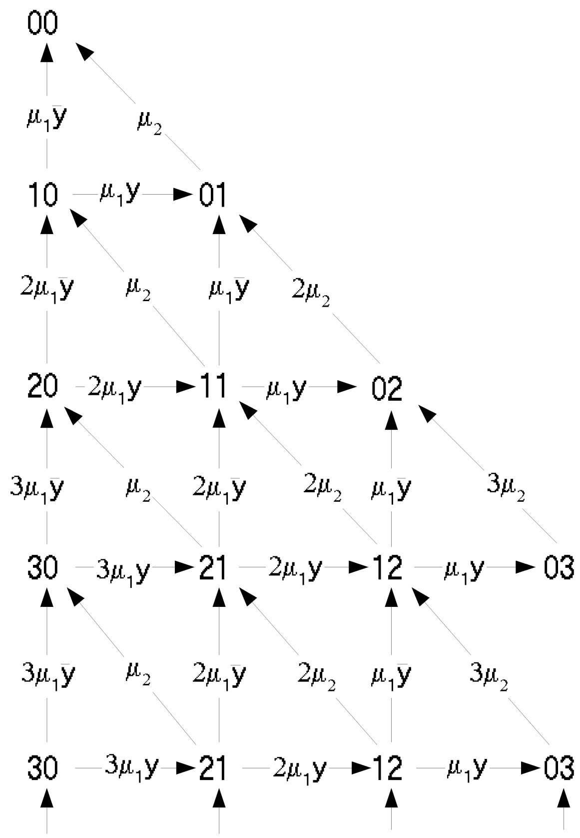

Рис. 2. Диаграмма переходов в системе *M/C*$_2$/3 по завершению обслуживания

Последовательные ярусы этой диаграммы соответствуют возрастающему числу заявок в системе (верхний – нулю). При числе заявок более трех (число каналов) остальные заявки ждут очереди, и набор «ключей» микросостояний, задающих расстановку обслуживаемых заявок по фазам обслуживания, меняться перестает.

Покажем далее использование гамма-распределения. В этом случае интегралы

$$q_j = \int_0^\infty \frac{(\lambda t)^j}{j!} e^{-\lambda t} \frac{\mu(\mu t)^{r-1}}{\Gamma(r)} e^{-\mu t}\, dt = \frac{\lambda^j \mu^r}{j!\Gamma(r)} \int_0^\infty t^{r+j-1} e^{-(\lambda+\mu)t}\, dt$$

$$= \frac{\lambda^j \mu^r}{j!(\lambda+\mu)^{r+j}\Gamma(r)} \int_0^\infty u^{r+j-1} e^{-u}\, du = \left(\frac{\mu}{\lambda+\mu}\right)^r \left(\frac{\lambda}{\lambda+\mu}\right)^j \frac{\Gamma(r+j)}{j!\Gamma(r)}, \quad j = 0,1,...$$

элементарно вычисляются рекуррентно. Прежде всего, $q_0 = (\mu/(\lambda+\mu))^r$. Заметим, что это выражение можно рассматривать как ПЛС от распределения, аппроксимированного гамма-плотностью. Далее,

$$\frac{q_j}{q_{j-1}} = \frac{\lambda}{\lambda+\mu} \frac{\Gamma(r+j)\cdot(j-1)!}{j!\Gamma(r+j-1)} = \frac{\lambda}{\lambda+\mu} \frac{r+j-1}{j}.$$

Таким образом, здесь даже не приходится вычислять гамма-функцию.

Законы сохранения

Законы сохранения позволяют быстро и компактно получать соотношения между основными параметрами и переменными задачи. В частности, с их помощью можно записать векторно-матричные уравнения баланса между вероятностями микросостояний систем обслуживания, марковизированных методом фиктивных фаз.

Обозначим через S_j множество всех возможных микросостояний системы, при которых на обслуживании находится ровно j заявок, а через σ_j – количество элементов в S_j. Далее в соответствии с диаграммой переходов для выбранной модели построим матрицы интенсивностей элементарных переходов:

$A_j\left[\sigma_j \times \sigma_{j+1}\right]$ – в S_{j+1} (прибытие заявки),

$C_j\left[\sigma_j \times \sigma_j\right]$ – в S_j (конец промежуточной фазы обслуживания),

$B_j\left[\sigma_j \times \sigma_{j-1}\right]$ – в S_{j-1} (полное завершение обслуживания заявки),

$D_j\left[\sigma_j \times \sigma_j\right]$ – в S_j (уход из состояний яруса j)

(в квадратных скобках указан размер матриц).

Введем векторы-строки $\gamma_j = \left\{\gamma_{j,1}, \gamma_{j,2}, \ldots, \gamma_{j,\sigma_j}\right\}$ нахождения системы в микросостояниях (j,i), $j = 0,1,\ldots,$ $i = 1,\ldots,\sigma_j$. Теперь можно записать векторно-матричные уравнения баланса переходов между состояниями

$$\gamma_0 D_0 = \gamma_0 C_0 + \gamma_1 B_1,$$

$$\gamma_j D_j = \gamma_{j-1} A_{j-1} + \gamma_j C_j + \gamma_{j+1} B_{j+1}, \quad j = 1,2,\ldots$$

Необходимые матрицы интенсивностей переходов между микросостояниями смежных ярусов можно построить автоматически – сопоставлением «ключей», задающих расстановку заявок по фазам обслуживания и/или прибытия для исходного и конечного микросостояний, в соответствии с типом перехода.

Полученные уравнения могут быть решены итерационным методом Такахаси – Таками или более быстрым, но менее универсальным методом матрично-геометрической прогрессии. По найденным вероятностям с помощью закона сохранения стационарной очереди можно рассчитать моменты распределения времени ожидания. Далее при необходимости выполняется их свертка в моментах с распределением чистой длительности обслуживания. По результирующим моментам строится ДФР длительности пребывания в системе.

Моменты и преобразование Лапласа

Решение многих задач теории приоритетных систем удается довести только до преобразований Лапласа – Стилтьеса (ПЛС) искомых распределений через ПЛС от известных. Поскольку последние можно вычислить с помощью уже обсуждавшейся технологии (по гамма-аппроксимации), открывается возможность получить таблицу значений искомого ПЛС в окрестности нулевого аргумента, после чего многократным численным дифференцированием ее (точнее, построенного по ней интерполяционного многочлена Ньютона или Стирлинга) получить искомые моменты.

Потокоэквивалентная декомпозиция сетей

Потокоэквивалентная декомпозиция сетей обслуживания предполагает расчет сети обслуживания в три этапа:

- составляется и решается система баланса средних интенсивностей входящих и выходящих потоков для узлов сети;

- каждый из узлов рассчитывается как изолированная система с найденной на первом этапе интенсивностью входящего потока. При необходимости выполняется уточнение *распределения* интервалов между входящими заявками с учетом операций суммирования (на входе узла), преобразования распределения от входа к выходу узла и прореживания на выходе – при выборе следующего узла. При наличии циклических маршрутов уточнение оказывается итерационным, но трудоемкий расчет каждого из узлов с рекуррентным входящим потоком приходится делать не более трех раз. По найденным распределениям заявок строится диагональная матрица ПЛС распределений времени пребывания *в узлах* при однократном посещении последних;
- с помощью упомянутой матрицы и матрицы маршрутов находится ПЛС распределения времени прохождения заявки от источника к стоку сети и по уже описанной технологии численного дифференцирования определяются моменты этого распределения. Далее можно построить его ДФР.

Имеется обобщение этой схемы на случай потоков неоднородных заявок.

Пакет подпрограмм МОСТ

Описанные выше и многие другие методы по мере их разработки автором получали программную реализацию и постепенно сформировали *пакет прикладных программ* МОСТ (Массовое Обслуживание – Стационарные задачи). Все процедуры пакета тщательно тестировались одним или несколькими из следующих методов:

- сопоставление с эталоном (например, процедуры численного дифференцирования – на ПЛС показательного закона $\beta(s) = \mu/(\mu + s)$, с которой связаны моменты $k!/\mu^k, \quad k = 1, 2, \ldots$;
- решением взаимообратных задач (подбор параметров аппроксимации по заданным моментам и вычисление моментов известного распределения заданного класса, обращение обратной матрицы или вычисление ее произведения напрямую);
- решением одной и той же задачи (средней длины очереди) при исходных данных, допускающих различные аппроксимации распределений и, соответственно, расчет с помощью разных процедур.

Первая версия пакета для ЕС ЭВМ (язык ПЛ/1) в составе 83 процедур в 1987 г. была передана в Государственный Фонд алгоритмов и программ (г. Таллинн) и уже в следующем году эксплуатировалась более чем в 30 организациях. Последующая из 113 процедур по сию пору используется в учебном процессе ВКА им. А.Ф. Можайского на ЕС-1066.

В настоящее время МОСТ дополнительно реализован на Фортране 77 (точнее, Fortran Microsoft 5.x) и работает под управлением MS DOS. *Профессиональная* версия включает около 140 процедур и свыше 90 тестов (всего более 20 тыс. строк фортранного кода). Она предполагает, что пользователь знает, по крайней мере, основы упомянутого Фортрана и базовые понятия теории очередей. Такой клиент должен сам составить ведущую программу – в основном из вызовов процедур пакета. В типичном случае цепочка вызовов включает в себя:

- расчет параметров аппроксимирующих распределений;
- основную процедуру расчета вероятностей состояний;
- вычисление факториальных моментов этих распределений;
- переход к моментам распределения времени ожидания;
- их свертку с моментами распределения чистой длительности обслуживания;
- построение таблицы ДФР.

Ведущая программа должна включать описания используемых объектов и может быть дополнена нестандартными вычислениями, циклическим перебором параметров и т. п. Пользователю открыт доступ на внутренний уровень пакета – к служебным процедурам, облегчающим его творческую работу по созданию новых моделей и методов.

Версия *для начинающих* имеет более ограниченные возможности, но соответственно легче в использовании. Она в режиме диалога с пользователем принимает от него исходные данные (не менее двух моментов исходных распределений) и по ним автоматически формирует и запускает фортран-программу. Результаты записываются в файл в формате, удобном для систем автоматизированного построения графиков.

Обе версии пакета обеспечены кратким введением в теорию очередей, руководствами по применению пакета и руководствами к лабораторным работам со студентами на его основе. Из этих работ особенно поучительна «*Экспериментальная проверка законов сохранения на имитационной модели*». Здесь исполнители в дополнение к включенной в пакет стандартной модели *GI/G/n/R* по указанному варианту (число каналов и тип распределения обслуживания, коэффициент загрузки) рассчитывают необходимые числовые данные и пишут генераторы случайных величин, настроенные на тип и параметры распределения. Далее в главной программе записываются цепочка вызовов модели и вспомогательных программ, а также проверки относительной невязки:

- закона сохранения заявок $\sum_{i=0}^{n-1}(n-i)\,p_i = n-\lambda b_1$ (вероятности дает модель, а в правую часть подставляются исходные данные);
- законов сохранения стационарной очереди $q_{[k]} = \lambda^k w_k$, где факториальные моменты длины очереди $q_{[k]}$ считаются по ответному распределению вероятностей, а моменты распределения ожидания w_k получаются из модели непосредственно;
- аналога предыдущей проверки для моментов распределения полного числа заявок в системе и времени пребывания заявок в ней. Заметим, что эта проверка в зависимости от числа каналов и типа распределения обслуживания должна давать разные результаты.

Обсуждаемая работа при небольшом объеме программирования обеспечивает твердое усвоение базовых понятий теории распределений, теории очередей и имитационного моделирования (типовые распределения, коэффициент загрузки, законы сохранения, принципы имитации, генераторы псевдослучайных чисел, методы понижения дисперсии). В связи с последними отметим, что в *модели* сознательно не определяются моменты полной длительности пребывания заявки в системе – эти показатели можно оценить с меньшей дисперсией.

По теоретическим основам пакета и его программной реализации в разные годы (начиная с конца 1970-х) опубликованы около двух десятков статей и несколько учебников. Подготовлена (и набрана в TeX'е) рукопись – гибрид монографии, хрестоматии и учебника объемом в 360 с.

Литература

1. **Рыжиков Ю.И.** Теория очередей и управление запасами. – СПб.: Питер, 2001.– 376 с.