УДК: 519:85, 65:01

ББК: 22:18, 65:290

Математическое моделирование систем массового обслуживания: методика принятия управленческих решений с помощью модели

Глухова Наталья Владимировна

Кандидат биологических наук, доцент кафедры высшей математики,

ФГБОУ ВО «Ульяновский Государственный педагогический университет им. И.Н. Ульянова», Ульяновск, Россия

Аннотация. В работе рассматривается математическая модель Колмогорова-Эрланга для систем массового обслуживания с отказами и очередями. Приводится алгоритм построения системы дифференциальных уравнений по схеме переходов для данной модели и показано, какие практические выводы можно сделать при помощи данной модели. В частности, описана методика оценки рентабельности увеличения численности штата (числа каналов обслуживания) или его сокращения. Приведены конкретные примеры ситуаций, в которых изменение численности штата является выгодным или невыгодными. Описаны условия, в которых для произвольной системы массового обслуживания с очередью необходимо введение искусственных ограничений на очередь (система предварительной записи).

Ключевые слова: системы массового обслуживания, вероятность, пропускная способность, отказ, очередь.

В литературе можно встретить различные определения систем массового обслуживания (с.м.о.), но все они содержательно сходны. Например, с.м.о. называется система, состоящая из нескольких (возможно одного) пунктов, предназначенных для обслуживания клиентов, которые случайным образом обращаются в данную систему [11, с. 47], [4, с. 34]. Б.А. Горлач определяет с.м.о. как системы, в которых на входе располагаются запросы на обслуживание, на выходе же – оказанные услуги [9, с. 269]. Примерами с.м.о. являются магазины, справочные, рестораны, почтовые отделения, службы консультирования по телефону, мастерские – то есть любые системы, предлагающие некоторые услуги, обратиться за которыми клиенты могут в случайное время (разумеется, в рамках времени работы) [2, с. 132]. Анализу деятельности с.м.о. посвящено большое количество литературы, в основном по исследованию операций. Большая часть источников является, на наш взгляд, перегруженной теоретической математической информацией. Так в работе [5, с. 950] отмечается сложность для восприятия менеджерами абстрактных математических рассуждений при решении задач динамического программирования. Это же можно отнести и к задачам, касающимся анализа с.м.о. Тем не менее, вопросы исследования операция имеют самое широкое практическое применение в сфере не только менеджмента, но и социальной [6, с. 130] и гуманитарной [7, с. 215 – 217] деятельности, в которой работают специалисты в большинстве не имеющие глубокого математического образования, для которых работа с математическими объектами сопряжена с серьёзными трудностями [8, с. 133]. Поэтому важной задачей представляется максимальное упрощение методов решения задач, возникающих в процессе принятия решений. В то же время имеются и работы, впадающие в другую крайность. В них без каких-либо обоснований даются готовые формулы, по которым рассчитываются основные характеристики с.м.о. Формулы эти громоздки, часто содержат не совсем понятные по смыслу задачи обозначения, сложны для запоминания. В качестве примера можно указать формулу для расчёта вероятности простоя p_0, которая без вывода приводится в работах [1, с. 262], [9, с. 283] для простейшего случая с.м.о. с отказами в установившемся режиме:

$$p_0 = (1 + \frac{\lambda}{\mu} + \frac{\lambda^{n2}}{2!\mu^2} + \ldots + \frac{\lambda^n}{n!\mu^n})^{-1}.$$

Данная величина p_0 затем в этом же параграфе используется как составная часть ещё в пяти формулах. И это только самый простой случай – в с.м.о. с очередями формулы становятся еще более громоздкими. Кроме того, применение формулы без понимания того, откуда она получилась, приводит не только к недоверию к результатам расчётов, но и к ошибочному её применению в тех ситуациях, к которым она неприменима (не формируется представление о сфере применимости модели, как это можно видеть в работе [1, с. 254 – 266]).

В тоже время, можно решать задачи по оптимизации с.м.о. и без глубокого изучения математической теории Марковских процессов. В настоящей статье мы опишем, следуя в основном работе [2, с 126 – 150] с некоторыми упрощениями, схему вывода дифференциальных уравнений Колмогорова-Эрланга, укажем алгоритм составления уравнений без подробностей вывода, приведем примеры применения вероятностных характеристик с.м.о., работающих в стабильном режиме. При этом будет использоваться минимум математического аппарата, доступный даже школьникам – потребуется только умение решать простейшие системы линейных уравнений, а также знание понятий «производная» и «вероятность» и формулы суммы геометрической прогрессии. Кроме того, в большинстве работ по с.м.о. при анализе моделей ограничиваются расчётом вероятностных характеристик без указания конкретного практического применения этих результатов. В качестве исключений можно отметить работу [3, с. 119], в которой предлагается определить число каналов, требуемое для того, чтобы обслуживались не мене 80 % заявок (в указанной работе задача решается по готовым громоздким формулам при помощи программы Excel) , а также работу [2, с. 146], в которой делается намёк на оптимизационную задачу, которую в этой работе не решают, а лишь предлагают сформулировать и решить «неленивому и любопытному читателю», и имеется задача о сравнении двух способов организации железнодорожных касс [2, с. 153 – 155]. В настоящей работе мы постараемся сделать акцент именно на практической применимости результатов к принятию конкретных управленческих решений.

Будем называть процесс обращения посетителей *потоком заявок*, а среднее количество обращений от клиентов за определённое время называется *интенсивностью* этого потока (обозначим её λ, как это принято в теории марковских процессов [2, с. 118]). Например, если в с.м.о. в среднем обращается один посетитель за 3 минуты, то λ = 1/3 з/мин. Поток посетителей, покидающих систему после того, как их обслужили, также обладает интенсивностью. Среднее количество клиентов, обслуживаемых одним каналом за определённое время будем обозначать символом μ. Например, если среднее время обслуживания одного клиента (Т) составляет 5 минут, то скорость обслуживания составляет 1/5 заявки в минуту (μ = 0,2 з/мин). Будем также полагать, что оба потока, протекающие в системе, являются стационарными, ординарными и не имеют последействия (подробнее с этими понятиями можно ознакомится в работах [2 с. 118 – 120], [4, с. 35].

Модели с.м.о. распадаются на два больших класса – *с отказами* (например, системы обслуживания по телефону – если клиент звонит на линию, по которой в данный момент разговаривают, то клиент слышит короткие гудки и вынужден звонить повторно в другое время) и *с очередью* (например, в супермаркете, если все кассы заняты, то покупатель становится в очередь). Важно также различать модели по типу очереди, которые могут быть неограниченными или ограниченными.

Пусть имеется с.м.о. с отказами с n пунктами обслуживания (каналами). В таком случае, данная система может находиться в одном из n + 1 состояний: S_0– все свободны, S_1 – в системе находится один посетитель и один пункт занят, … , S_n – все пункты заняты. Здесь, для простоты, примем, что обслуживающий персонал во всех пунктах работает с одинаковой скоростью, поэтому важно только общее количество работающих пунктов, а не то, какие конкретно пункты работают. Например, если регистратура в поликлинике с тремя окошками обслуживает 2 пациентов, то важно, что одно окошко свободно; какое именно – неважно (это справедливо, если работники регистратуры работают одинаково; если один работает медленнее, чем другие то требуется более детальный анализ).

В начальный момент работы системы (когда она только открылась) в ней ещё нет заявок, следовательно, с.м.о. находится в состоянии S_0. При первом обращении с.м.о. переходит в состояние S_1. Скорость этого перехода определяется интенсивностью λ. Из состояния S_1 система может перейти как в состояние S_2 (с той же интенсивностью λ), так и в S_0 (заявка обслужена и система снова свободна, скорость данного перехода характеризуется величиной μ). Аналогично из S_2 с.м.о. может перейти в состояние S_3 (появилась третья заявка, три канала заняты, интенсивность λ), или в S_1 (переход от S_2 к S_1 осуществляется со скоростью 2μ, так как в состоянии S_2 уже работает два пункта, а значит и их общая скорость в два раза больше). Продолжая аналогичные рассуждения, можем составить следующую схему:

$$\boxed{S_0} \underset{\mu}{\overset{\lambda}{\rightleftarrows}} \boxed{S_1} \underset{2\mu}{\overset{\lambda}{\rightleftarrows}} \boxed{S_2} \underset{3\mu}{\overset{\lambda}{\rightleftarrows}} \;\ldots\; \underset{n\mu}{\overset{\lambda}{\rightleftarrows}} \boxed{S_n}$$

Обозначим через $p_k(t)$ вероятность того, что в момент времени t система находится в состоянии S_k. Опишем теперь, чему будет равна вероятность того, что система будет находиться в состоянии S_k через небольшой промежуток времени Δt. Вероятность каждого перехода из одного состояния в другое за указанный промежуток времени есть произведение длительности этого промежутка на интенсивность соответствующего данному переходу потока. В состояние S_k мы можем попасть одним из следующих трёх способов.

1. Из состояния S_{k-1}, если за время Δt к нам обратится еще один клиент – вероятность этого $\lambda\Delta t\, p_{k-1}(t)$.

2. Из состояния S_{k+1}, если за данный промежуток Δt одна клиент завершил обслуживание, то есть один пункт освободился – вероятность этого события равна $(k+1)\mu\Delta t p_{k+1}(t)$.

3. Из состояния S_k, если ничего нового не произошло – вероятность того, что с.м.о. осталась в S_k, составляет $(1 - \lambda\Delta t - k\mu\, \Delta t)p_k(t)$.

$$\ldots\; \boxed{S_{k-1}} \underset{k\mu}{\overset{\lambda}{\rightleftarrows}} \boxed{S_k} \underset{(k+1)\mu}{\overset{\lambda}{\rightleftarrows}} \boxed{S_{k+1}} \;\ldots$$

Суммируя все возможные варианты получаем:

$$p_k(t + \Delta t) = \lambda\Delta t\, p_{k-1}(t) + (k+1)\mu\Delta t\, p_{k+1}(t) + (1 - \lambda\Delta t - k\mu\, \Delta t)p_k(t).$$

Проведя очевидные преобразования и перейдя к пределу при $\Delta t \to 0$, получим:

$$p_k^{/}(t) = \lim_{\Delta t \to 0} \frac{p_k(t + \Delta t) - p_k(t)}{\Delta t} = \lambda p_{k-1}(t) + (k+1)\mu p_{k+1}(t) - \lambda p_k(t) - k\mu p_k(t).$$

Вывод этого дифференциального уравнения совершенно не требуется воспроизводить каждый раз. Достаточно применить следующий алгоритм:

1) записываем $p_k^{/}(t)$ для каждого состояния S_k;

2) рассматриваем последовательно все стрелки, соединённые с квадратом, изображающим данное состояние на графе;

3) если стрелка входит в квадрат, изображающий рассматриваемое состояние, то ставим знак плюс, а если выходит, то минус;

4) записывается интенсивность, стоящая рядом со стрелкой;

5) интенсивность умножается на вероятность того состояния, откуда стрелка выходит;

6) переходим к следующей стрелке, если ещё не все стрелки описаны (возврат к п. 3), либо переходим к составлению следующего уравнения (п. 1).

В полученную систему уравнений добавляем еще одно условие: сумма вероятностей всех возможных состояний должна быть равна единице:

$p_0(t) + p_1(t) + p_2(t) + \ldots + p_n(t) = 1.$

В результате получим систему, называемую *системой дифференциальных уравнений Колмогорова-Эрланга* (названа в честь её авторов).

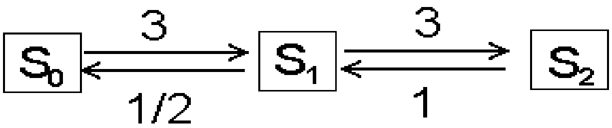
На начальном этапе работы, например, с.м.о. работает не стабильно, так как количество клиентов может существенно меняться – либо о с.м.о. еще не знают, поэтому клиентов мало, либо, напротив, была произведена рекламная акция, поэтому количество клиентов будет завышенным, но через какое-то время все процессы, как правило, стабилизируются. В практических задачах нас чаще всего интересуют именно характеристики такого стабильного (установившегося) режима [10, с. 612]. В этом случае, все вероятности $p_k(t)$ – это просто числа (будем обозначать их символами p_k). В теории случайных процессов доказывается утверждение «если число состояний системы конечно, и из каждого из них можно за конечное число шагов перейти в любое другое, то вероятности стабильного состояния существуют» [2, с. 129]. Если все $p_k(t)$ – константы, то их производные равны нулю, и уравнения из дифференциальных превращаются в линейные:

$$\begin{cases} \mu p_1 - \lambda p_0 = 0 \\ \lambda p_0 - \mu p_1 - \lambda p_1 + 2\mu p_2 = 0 \\ \lambda p_1 - 2\mu p_2 - \lambda p_2 + 3\mu p_3 = 0 \\ \ldots \\ \lambda p_{n-1} - n\mu p_n = 0 \\ p_0 + p_1 + \ldots + p_n = 1 \end{cases},$$

Полученную систему уже легко решить без специальных математических знаний.

Пример 1: Имеется коммутатор с двумя параллельными телефонами, на которых работают одновременно две телефонистки. Клиенты звонят на коммутатор в среднем 3 раза каждую минуту, среднее время разговора – 2 минуты. Пусть среднее число обращений на коммутатор не зависит от времени суток (поток заявок стационарен). Найдите вероятности с.м.о, работающей в стабильном режиме.

<u>Решение</u>: Составим схему переходов для данной системы. Количество пунктов обслуживания n = 2, интенсивность обращений λ = 3 з/мин, интенсивность обслуживаний μ = ½ з/мин:



По указанному выше алгоритму составляем систему дифференциальных уравнений:

$$\begin{cases} p_0^{/}(t) = -3p_0(t) + \frac{1}{2}p_1(t) \\ p_1^{/}(t) = 3p_0(t) - \frac{1}{2}p_1(t) - 3p_1(t) + p_2(t) \\ p_1^{/}(t) = 3p_1(t) - p_2(t) \\ p_0(t) + p_1(t) + p_2(t) = 1 \end{cases}.$$

Если система стабилизировалась, то все производные оказываются равными нулю, все функции вероятностей становятся константами, то есть:

$$\begin{cases} -3p_0 + \dfrac{1}{2}p_1 = 0 \\ 3p_0 - \dfrac{1}{2}p_1 - 3p_1 + p_2 = 0 \\ 3p_1 - p_2 = 0 \\ p_0 + p_1 + p_2 = 1 \end{cases}.$$

Выражаем из первого равенства p_1 через p_0, затем складываем первое уравнение со вторым и выражаем p_2, подставляем результаты в последнее:

$$\begin{cases} p_1 = 6p_0 \\ p_2 = 3p_1 = 18p_0 \\ p_0 + 6p_0 + 18p_0 = 1 \end{cases}.$$

Тогда $25p_0 = 1$ откуда следует, что $p_0 = 1/25 = 0{,}04$, $p_1 = 6p_0 = 0{,}24$, $p_2 = 0{,}72$. Таким образом, в установившемся режиме с.м.о. будет полностью свободна в 4 % случаев, в 24 % случаев будет работать одна телефонистка, а в 72 % случаев они будут работать обе.

Для практических нужд полезно определить следующие величины.

Вероятность простоя – вероятность, что клиентов в системе нет, это p_0.

Вероятность отказа – вероятность того, что клиент не будет обслужен по причине занятости всех пунктов обслуживания (она равна вероятности последнего состояния на схеме, p_n).

Относительная пропускная способность (Q) – процент или доля клиентов из числа обратившихся, которые получат обслуживание. Вычисляется вычитанием из 1 вероятности отказа. В примере 1 в 72 % случаев все телефонистки будут заняты, следовательно, 72 % позвонивших получат отказ. Значит, будет обслужено только 28 % позвонивших (Q= 0,28).

Абсолютная пропускная способность (А) – реальное количество клиентов, которые с.м.о. обслужит за заданный период времени. Её можно рассчитать, вычислив указанный процент обслуженных заявок (Q) от числа обратившихся (λ), то есть $A = \lambda Q$. В приведённом выше примере $A = 3 \cdot 0{,}28 = 0{,}84$

з/мин. Найденную величину A можно перевести в другие единицы времени. Поскольку все интенсивности рассчитывались за минуту, то и абсолютная пропускная способность будет получена за минуту. Но мы можем подсчитать, сколько заявок обслуживает наша система за час, умножив данный результат на 60. Для данного примера A = 50,4 заявки в час. Умножив 50,4 на количество рабочих часов в день, получим суточную пропускную способность системы и т.д.

Знание абсолютной пропускной способности позволяет спрогнозировать рентабельность процедур сокращения или увеличения числа обслуживающих каналов. Если система, как в нашем примере, перегружена, то есть обслуживается менее 30 % клиентов, то можно ожидать, что увеличение количества каналов может привести к увеличению прибыли, так как большее число клиентов будет обслужено. Однако, это не всегда справедливо, так как содержание дополнительного канала всегда связано с расходами – нужно выплачивать зарплату дополнительной телефонистке, а также оплачивать телефонную линию, а возможно и дополнительное помещение. Зная абсолютную пропускную способность, а также доход, получаемый от обслуживания одного звонка, мы можем сравнить доходы за определенный период времени (которые мы вычислим, перемножив две эти величины) с расходами и понять, будет ли данная процедура рентабельна или нет. Также можно рассмотреть и вопрос о рентабельности сокращения штата.

Решим задачу в следующих условиях. Пусть каждый клиент за обслуженный звонок платит коммутатору 2 рубля. Содержание одной линии коммутатора (оплата телефонной связи и зарплата телефонисткам) обходится в 500 рублей в сутки, плюс 200 рублей в сутки за арендуемое помещение, в котором сидят обе работающие телефонистки. Помещение позволяет разместить в нем еще одну линию. Коммутатор работает круглосуточно. Требуется выяснить, будет ли выгодным подключение дополнительной линии. При двух каналах коммутатор получает $24 \cdot 50{,}4 = 1209{,}6$ звонков в сутки и, следовательно, клиенты платят в среднем 1209,6·2 = 2419,2 руб. каждый день. Расходы составляют 1200 руб. в день, то есть чистая прибыль, которую приносит коммутатор в настоящее время, составляет 1219,2 руб. в сутки. Рассмотрим, что произойдёт, если мы подключим третью линию. Схема переходов для такой ситуации будет иметь вид:

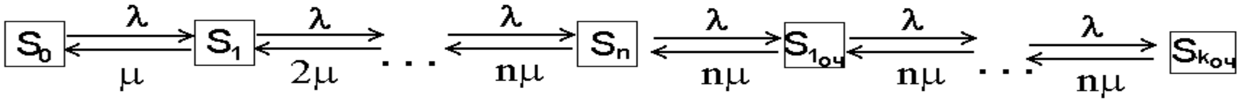
$$\boxed{S_0} \underset{1/2}{\overset{3}{\rightleftarrows}} \boxed{S_1} \underset{1}{\overset{3}{\rightleftarrows}} \boxed{S_2} \underset{3/2}{\overset{3}{\rightleftarrows}} \boxed{S_3}$$

Откуда получим систему:

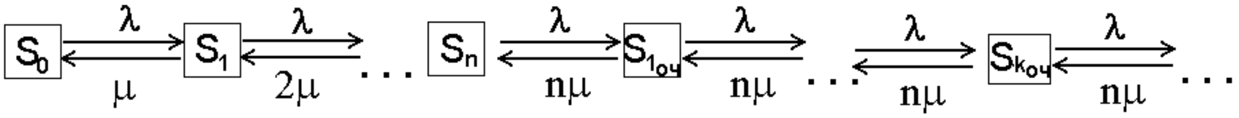
$$\begin{cases} -3p_0 + \frac{1}{2} p_1 = 0 \\ 3p_0 - \frac{1}{2} p_1 - 3p_1 + p_2 = 0 \\ 3p_1 - p_2 - 3p_2 + \frac{3}{2} p_3 = 0 \\ 3p_2 - \frac{3}{2} p_3 = 0 \\ p_0 + p_1 + p_2 + p_3 = 1 \end{cases}$$

Здесь, как и ранее, $p_1 = 6p_0$, $p_2 = 18p_0$, и кроме того появляется $p_3 = 2p_2 = 36p_0$. Тогда $p_0 + p_1 + p_2 + p_3 = 61p_0$, откуда $p_0 = 1/61$, $p_{отк} = p_3 = 36/61$, $Q = 25/61$. Умножая Q на $\lambda = 3$, получаем, что увеличенный коммутатор сможет обслужить 75/61 заявок за минуту или $24 \cdot 60 \cdot \frac{75}{61} = \frac{108000}{61}$ заявки в сутки. Это приносит доход $\frac{216000}{61} \approx 3540$ руб. 98 коп. Расходы на содержание коммутатора при этом также возрастут, так как появится дополнительная линия, и составят 1700 руб. в сутки. Тогда чистая среднесуточная прибыль будет равна 1840 руб. 98 коп, что более чем на 600 рублей превысит прибыль от коммутатора с двумя линиями. Поэтому следует принять решение об увеличении количества каналов.

Анализ работы с.м.о. с очередью, ограниченной k местами очень похож на анализ с.м.о. с отказами, если только в схему состояний добавить состояния $S_{1оч}$ – в с.м.о. одна заявка ждёт своей очереди, $S_{2оч}$ – две заявки в очереди, $S_{kоч}$ – в с.м.о. k заявок в очереди. Схема переходов для этой с.м.о. имеет вид:



Ситуация резко осложняется, если очередь не ограничена, то есть количество возможных состояний бесконечно: мы не можем определить максимально возможное количество человек в очереди. Поэтому нам придётся иметь дело с системой, содержащей бесконечное число уравнений с бесконечным числом неизвестных. Однако и из этой ситуации есть выход. Запишем систему переходов и систему уравнений для с.м.о. с неограниченной очередью в установившемся режиме:



$$\begin{cases} \mu p_1 - \lambda p_0 = 0 \\ \lambda p_0 - \mu p_1 - \lambda p_1 + 2\mu p_2 = 0 \\ \ldots \\ \lambda p_{n-1} - n\mu p_n - \lambda p_n + n\mu p_{1оч} = 0 \\ \lambda p_n - n\mu p_{1оч} - \lambda p_{1оч} - n\mu p_{2оч} = 0 \\ \ldots. \\ p_0 + p_1 + \ldots p_n + p_{1оч} + p_{2оч} + \ldots = 1 \end{cases}$$

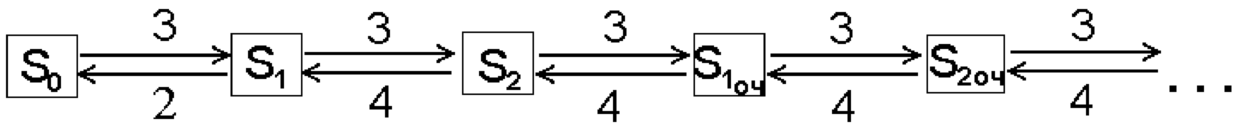
Последовательно складывая уравнения и выражая p_k через p_0, получим:

$$\begin{cases} p_1 = \frac{\lambda}{\mu} p_0 \\ p_2 = \frac{\lambda}{2\mu} p_1 = \frac{\lambda^2}{2\mu^2} p_0 \\ \ldots \\ p_n = \frac{\lambda}{n\mu} p_{n-1} \\ p_{1оч} = \frac{\lambda}{n\mu} p_{1оч} \\ \ldots \\ p_0 + p_1 + \ldots p_n + p_{1оч} + p_{2оч} + \ldots = 1 \end{cases}$$

Можно заметить, что, начиная с p_n, каждая следующая переменная получается умножением предыдущей на выражение $\frac{\lambda}{n\mu}$, то есть в последнем уравнении системы имеем, начиная с p_n, геометрическую прогрессию. Если $\frac{\lambda}{n\mu} \geq 1$, то бесконечная сумма будет возрастать до бесконечности. Это означает, что очередь будет неограниченно возрастать (система не будет справляться с поступающим потоком заявок). Поэтому, прежде чем составлять систему дифференциальных уравнений, полезно проверить данное условие. Если это так, то необходимо увеличивать количество каналов, либо, если это невозможно по тем или иным соображениям вводить искусственное ограничение на очередь. Отметим, что если система не является монополистом, то реально очередь бесконечно возрастать не будет – клиенты просто будут уходить из этой с.м.о. и обращаться в другие, так как очередь слишком велика. Поэтому данная с.м.о. может потерпеть очень большие убытки от того, что у неё слишком мало персонала. Во всяком случае, никакие прогнозы, построенные исходя из предположения бесконечной очереди, не оправдаются, наиболее вероятным будет постепенное уменьшение интенсивности клиентов до уровня, когда это условие перестанет выполняться. Если система монополист (например, пенсионный фонд или единственная поликлиника в населенном пункте), то людям просто некуда будет деваться, и они все равно будут создавать эту очередь. В такой ситуации необходимо ввести искусственное ограничение на очередь (предварительная запись, либо талоны на приём), а далее решать задачу с ограниченной очередью. Если $\frac{\lambda}{n\mu} < 1$, то геометрическая прогрессия – бесконечно убывающая, её сумму можно вычислить по формуле: $S = \frac{b_0}{1-q}$, где b_0 первый член данной прогрессии (в нашем случае $b_0 = p_n$), а q – знаменатель геометрической прогрессии ($q = \frac{\lambda}{n\mu}$). Применив данную формулу можно вычислить p_0, а значит и любые другие вероятности.

Пример 2. В супермаркет с 2 кассами обращается в среднем 3 покупателя в минуту, один кассир в среднем успевает обслужить 2 покупателя в минуту. Найти вероятность простоя данной системы.

Решение: Интенсивность потока заявок $\lambda = 3$ з/мин, интенсивность потока обслуживаний $\mu = 2$ з/мин. Схема переходов имеет вид:



В установившемся режиме получим систему уравнений:

$$\begin{cases} -3p_0 + 2p_1 = 0 \\ 3p_0 - 2p_1 - 3p_1 + 4p_2 = 0 \\ 3p_1 - 4p_2 - 3p_2 + 4p_{1оч} = 0 \\ 3p_2 - 4p_{1оч} - 3p_{1оч} + 4p_{2оч} = 0 \\ \ldots \\ p_0 + p_1 + p_2 + p_{1оч} + p_{2оч} + \ldots = 1 \end{cases}$$

откуда

$$\begin{cases} p_1 = \frac{3}{2}p_0 \\ p_2 = \frac{3}{4}p_1 = \frac{9}{8}p_0 \\ p_{1оч} = \frac{3}{4}p_2 = \frac{3}{4}\cdot\frac{9}{8}p_0 \\ p_{2оч} = \frac{3}{4}p_{1оч} = \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{9}{8}p_0 \\ \ldots \\ p_0 + \frac{3}{2}p_0 + \frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{9}{8}p_0 + \ldots = 1 \end{cases}$$

Вычислим сумму геометрической прогрессии, начинающейся с $p_2 = (9/8)p_0$

$$\frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{9}{8}p_0 + \ldots = \frac{\frac{9}{8}p_0}{1-\frac{3}{4}} = \frac{\frac{9}{8}p_0}{\frac{1}{4}} = \frac{9}{2}p_0 .$$

Тогда $p_0 + \frac{3}{2}p_0 + \frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{9}{8}p_0 + \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{9}{8}p_0 + \ldots = p_0 + \frac{3}{2}p_0 + \frac{9}{2}p_0 = 7p_0 = 1,$

откуда, $p_0 = 1/7$ (вероятность простоя).

При анализе с.м.о. с очередью полезно знать ещё две характеристики этой системы: среднюю длину очереди ($L_{оч}$) и среднее время, проводимое заявкой в очереди ($W_{оч}$). Следует учитывать, что если эти величины окажутся неразумно большими, то, например, покупатели в магазине, видя такую большую очередь, скорее всего не будут дожидаться обслуживания и будут покидать нашу систему, не дожидаясь обслуживания, то есть поток заявок перестанет быть простейшим (нарушается условие отсутствия последействия). В таких случаях анализ, проведённый по описанным выше правилам, станет некорректным и приведёт нас к ошибочным результатам (например, мы получим завышенные значения прибыли, не соответствующие действительности, так как часть клиентов мы просто потеряем). Для оценки $L_{оч}$, умножим количество людей в очереди в каждом состоянии на вероятность этого состояния:

$$L_{оч} = 1 \cdot p_{1оч} + 2 \cdot p_{2оч} + 3 \cdot p_{3оч} + \ldots$$

Здесь мы опять имеем дело с бесконечным числом слагаемых, которые, к сожалению, уже не образуют геометрическую прогрессию. Однако, эту сумму можно разложить на сумму нескольких геометрических прогрессий, поменяв местами слагаемые (вообще говоря в бесконечных суммах это надо делать с некоторой осторожностью, но данное действие является допустимым, если задача имеет решение, то есть, если $q = \dfrac{\lambda}{n\mu} < 1$):

$$L_{оч} = p_{1оч} + p_{2оч} + p_{2оч} + p_{3оч} + p_{3оч} + p_{3оч} \ldots =$$

$$= (p_{1оч} + p_{2оч} + p_{3оч} + \ldots) + (p_{2оч} + p_{3оч} + \ldots) + (p_{3оч} + p_{4оч} + \ldots).$$

Выражение в каждой скобке есть сумма геометрической прогрессии, то есть:

$$L_{оч} = \frac{p_{1оч}}{1-q} + \frac{p_{2оч}}{1-q} + \frac{p_{3оч}}{1-q} + \ldots = \frac{p_{1оч}}{1-q} + \frac{p_{1оч}q}{1-q} + \frac{p_{1оч}q^2}{1-q} + \ldots = \frac{p_{1оч} + p_{1оч}q + p_{1оч}q^2 + \ldots}{1-q}.$$

В числителе мы вновь получаем бесконечно убывающую геометрическую прогрессию с тем же знаменателем. Таким образом,

$$L_{оч} = \frac{\frac{p_{1оч}}{1-q}}{1-q} = \frac{p_{1оч}}{(1-q)^2}.$$

Среднее время, проводимое заявкой в очереди ($W_{оч}$), можно определить пользуясь следующим рассуждением. Если нам известно из каких-либо соображений, что в очереди мы будем стоять $W_{оч}$ минут, то к тому моменту, когда подойдёт наша очередь, за нами будет стоять $\lambda W_{оч}$ человек (все пришедшие после нас будут за нами), то есть $L_{оч} = \lambda W_{оч}$, откуда получаем формулу:

$$W_{оч} = \frac{L_{оч}}{\lambda}.$$

Рассчитаем $L_{ср}$ и $W_{оч}$ для с.м.о. из примера 2. Так как $p_{1оч} = \frac{3}{4} \cdot \frac{9}{8} p_0$, а $p_0 = 1/7$, значит $p_{1оч} = \frac{27}{224}$,

$$L_{оч} = \frac{p_{1оч}}{(1-\frac{\lambda}{n\mu})^2} = \frac{\frac{27}{224}}{(1-\frac{3}{4})^2} = \frac{\frac{27}{224}}{\frac{1}{16}} = \frac{27}{14},$$

$$W_{оч} = \frac{1}{3} \cdot \frac{27}{14} = \frac{9}{14} \text{ мин.}$$

Видим, что в среднем в нашей очереди будет стоять 1 – 2 человека, и время, проводимое ими в очереди будет составлять менее одной минуты, что вполне приемлемо, а, следовательно, нет необходимости увеличивать количество каналов обслуживания или вводить искусственные ограничения на очередь. При этом сокращение числа каналов до одного приведёт к нарушению условия $\frac{\lambda}{n\mu} < 1$, то есть сокращение персонала для данной с.м.о. также недопустимо, принимаем решение о том, что количество персонала остаётся неизменным.

Литература

1. Афанасьев М.Ю., Багриновский К.А., Матюшок В.М. Прикладные задачи исследования операций: Учебное пособие. М.: Инфра-М, 2006. 352 с.
2. Вентцель Е.С. Исследование операций: задачи, принципы, методология: учеб. пособие для втузов. М.: Высшая школа, 2001. 206 с.
3. Верник А.Н., Эткин А.Е., Эткина Г.П. Математические методы и модели в экономике. Ульяновск, УлГТУ, 2008. 208 с.
4. Глухова Н.В. Методы оптимизации использования трудовых ресурсов. Учебное пособие. Ульяновск, ФГБОУ ВО УлГПУ, 2017. 50 с.
5. Глухова Н.В. Новая методика изучения темы «динамическое программирование» на примере задачи об инвестировании для студентов, обучаю-

щихся экономике и управлению // Фундаментальные исследования. – 2014. № 8-4. С. 950–954.

6. Глухова Н.В. О мотивации изучения математических дисциплин студентами, обучающимися по направлению подготовки «Социальная работа» // Проблемы современного математического образования в высшей школе: Материалы международной заочной научной конференции. Ульяновск: УлГПУ, 2013. С. 130–134.
7. Глухова Н.В., Гришина С.А. Менеджериальные задачи как средство гуманитаризации математического образования // Гуманизация и гуманитаризация образования XXI века. Проблемы современного образования. Материалы 12-ой межд. науч.-метод. конф. памяти И.Н. Ульянова (19–20 октября 2011 г.). Ульяновск, УлГПУ, 2011. С. 215–217.
8. Глухова Н.В., Фолиадова Е.В. О применении задач с междисциплинарным содержанием при проведении олимпиад среди школьников и студентов // Актуальные вопросы методики обучения математике и информатике в условиях стандартизации образования. Материалы Всерос. науч.-практ. конф. препод. мат., информ. школ и вузов. Ульяновск: УлГПУ, 2016. С. 133–139.
9. Горлач Б.А. Исследование операций: учебный комплекс для студентов вузов, обучающихся по экономическим и техническим специальностям. Самара: Аэропринт, 2008. 370 с.
10. Таха Х.А. Введение в исследование операций. М.: Вильямс, 2001. 912 с.
11. Теория вероятностей с элементами математической статистики и анализа систем массового обслуживания / сост. Н.А. Волкова, Н.В. Глухова. Ульяновск, 2009. Т. 2, ч. 2. 76 с.