МОДЕЛЬ РЕШЕНИЯ ЗАДАЧ В АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ

А. Ю. Иванов, кандидат технических наук, доцент.
Санкт-Петербургский университет ГПС МЧС России

Разработана модель обработки запросов к распределению баз данных для автоматизированных информационных систем. Использован агрегативный подход для разработки систем; в качестве агрегатов применены объекты, с помощью которых генерируются и преобразуются моделируемые процессы. Показана динамическая связь между этими агрегатами. Разработана структурная схема модели.

Ключевые слова: модель решения задач, автоматизированные информационные системы, функция, агрегативный подход

В настоящее время интенсивно проводятся работы в сфере комплексной информатизации управления в МЧС. Основу процессов информатизации составят автоматизированные информационные системы (АИС). Цель внедрения этих систем в практику управления состоит в повышении оперативности и обоснованности принимаемых решений за счет более полного использования информации. Накопление, хранение и выдачу информации для решения задач управления в автоматизированном режиме обеспечивают базы данных. Территориальная рассредоточенность элементов системы управления МЧС определяет необходимость построения и использования распределенных баз данных (РБД). Очевидно, что временные показатели процессов решения задач в существенной степени зависят от характеристик РБД. В связи с этим оценка оперативности АИС и своевременности информационного процесса в системе управления требует построения и использования двухуровневого комплекса математических моделей. На нижнем уровне анализу подлежит процесс обработки запросов к РБД с целью оценки реактивности базы данных. Результаты этой оценки должны выступать в качестве исходных данных для получения интегрированных характеристик на верхнем уровне – при получении вероятностно-временных показателей процесса решения задач управления в АИС.

Модель обработки запросов к распределенной базе данных

По характеру функционирования АИС относятся к стохастическим сетям массового обслуживания (СеМО). Сеть представляет собой совокупность взаимосвязанных систем массового обслуживания (СМО), каждая из которых интерпретирует один из элементов АИС.

Основным показателем оперативности функционирования подобных сетей является функция распределения времени между поступлением и удовлетворением запроса и/или его моменты [1].

Стохастическая СеМО, определяется следующей совокупностью характеристик [2–4]:

1) множеством СМО $\{S_1,S_2,...,S_n\}$, образующих сеть;

2) числом каналов $K_1,K_2,..,K_n$ в системах $S_1,S_2,...,S_n$, соответственно;

3) матрицей траекторий движения заявок $R=||r_{ij}||$, где r_{ij} – номер СМО, на которую переходит заявка, продвигающаяся по *i*-му пути на *j*-й фазе обслуживания при детерминированной процедуре маршрутизации, или матрицей вероятностей перехода заявок из одной СМО в другую $P =||p_{ij}||$, где p_{ij} – вероятность того, что заявка, покидающая S_i, поступает в S_j;

4) числом заявок, циркулирующих в замкнутой сети (Z);

5) интенсивностью источников заявок в разомкнутой сети $\Lambda=\{\lambda_j^{(k)}\}$, где j – тип заявки; $к$ – категория срочности;

6) законами распределения времени $F_1^{(k)}(t)$, $F_2^{(k)}(t),\ldots,F_n^{(k)}(t)$ и дисциплинами обслуживания заявок в системах S_1, S_2,.., S_n.

Элементами, участвующими в обработке запросов к РБД и представляющими различные СМО, являются:

1) сети передачи данных (моноканалы – МК) локальных вычислительных сетей, входящих в состав общей сети АИС;

2) серверы базы данных (СБД), на которых размещены ФД;

3) коммутационные устройства (КУ), обеспечивающие обмен данными между удаленными узлами;

4) каналы передачи данных (КПД).

Помимо числа обслуживающих приборов, каждый тип СМО характеризуется следующими параметрами:

1) интенсивностью обслуживания заявок $\mu_{ij}^{(k)}$, где i – тип системы, j – тип заявки, k – ее категория срочности;

2) длиной очереди (если таковая имеется);

3) интенсивностями входного $\lambda_{вх\mathrm{i}}^{(\mathrm{k})}$ и выходного $\lambda_{вых\mathrm{i}}^{(\mathrm{k})}$ потоков заявок;

4) механизмом дообслуживания прерванных заявок

Заявки описываются вектором $\bar{v}=(t,k,j,d,a_1.a_2)$, где t – время поступления заявки в СеМО, k – категория срочности, j – тип, d – длина, a_1 и a_2 – адреса отправителя и получателя соответственно.

Расчет показателей реактивности РБД сводится к отысканию функции распределения времени пребывания заявок в СеМО. Как показано в теории распределений [5, 6], для удовлетворительного описания функции распределения случайной величины достаточно найти первые четыре центральные момента этой величины и при необходимости аппроксимировать закон распределения с помощью семейств кривых.

При построении модели целесообразно применить агрегативный подход. В качестве агрегатов выступают объекты, с помощью которых генерируются и преобразуются моделируемые процессы. Для моделирования процесса обработки запросов к РБД, представленной в виде СеМО, в качестве агрегатов выступают: источники заявок, системы массового обслуживания и управляющие процедуры.

Учитывая конкретную предметную область, можно выделить конечное множество типовых агрегатов – СМО, интерпретирующих элементы АИС, а также источники заявок в соответствии с типами запросов, поступающих в систему. По своей сути эти агрегаты представляют собой модели источников запросов и элементов АИС.

Динамические связи между этими агрегатами осуществляются с помощью процедуры, моделирующей процесс движения заявок и выполняющей функции коммутации входов и выходов агрегатов. Структурная схема модели, представленной с помощью агрегативного подхода, показана на рис. 1.

В качестве выходных переменных (результатов моделирования) могут выступать значения показателей реактивности РБД (среднее время отклика на запросы), а также дополнительные характеристики, необходимые для работы модели верхнего уровня. К ним относятся функции распределения и моменты времени пребывания в сети заявок каждого типа и категории срочности.

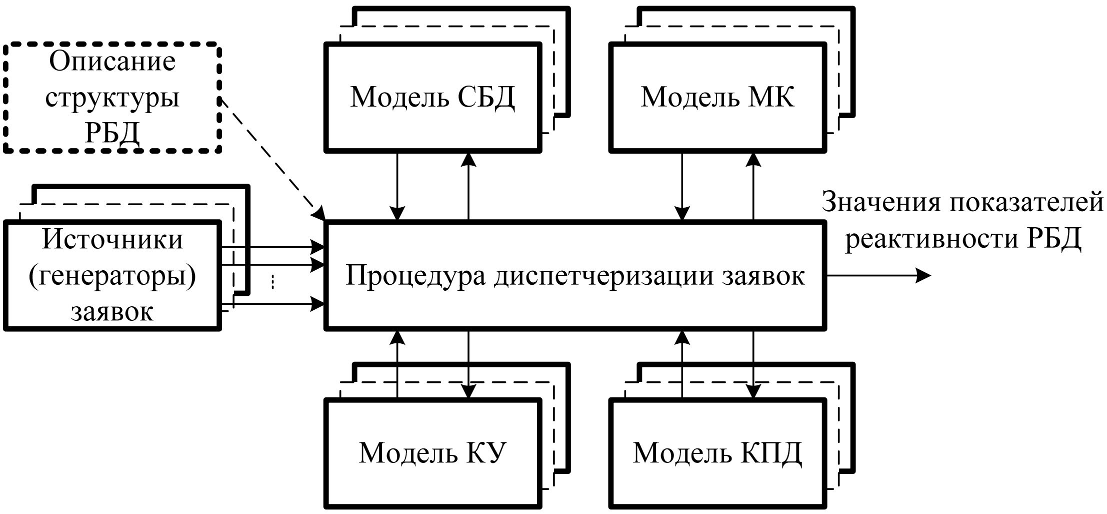

Рис. 1. **Структурная схема модели процесса обработки запросов к РБД**

Модель решения задач в АИС

В модели верхнего уровня процесс решения задачи в АИС представляется ориентированным графом, узлы которого отображают операции по обработке запросов к РБД, а дуги – связи между операциями (рис. 2)

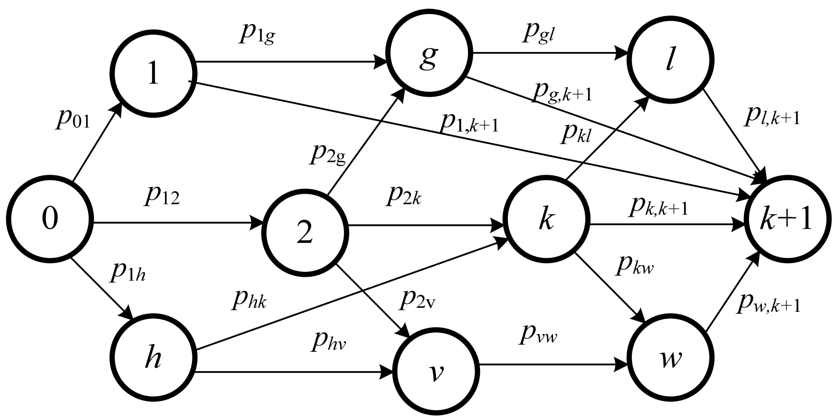

Рис. 2. **Граф процесса решения задачи**

Узел 0 является начальным, а узел k+1 – конечным. Каждому узлу ставится в соответствие вектор, определяющий параметры закона распределения времени его обработки. Дуги характеризуются вероятностями p_{kl} перехода от k-го запроса к l-му ($k=\overline{0,K}; l=\overline{1,K+1}$). Если из узла выходит единственная дуга, то переход по ней идет с вероятностью, равной 1. Если из узла выходит несколько дуг (k,l), (k,m),…,(k,w), то выбор направления рассматривается как случайное событие, характеризуемое вероятностями исходов p_{kl}, p_{km},…,p_{kw}, причем $\sum_{j=l}^{w} p_{kj} = 1$.

Определение вероятностно-временных показателей решения задач в АИС состоит в нахождении значений функции распределения времени по характеристикам элементов модели.

Исследование представленной модели возможно с помощью аппарата марковских цепей [7]. При этом задача сводится к расчету числовых характеристик многомерной случайной величины $X=(x_1, x_2,\dots,x_M)$ – количества попаданий случайного процесса в состояния $s_1, s_2,\dots,s_M$, соответствующие выполнению запросов. На основе полученных значений и и заданных характеристик узлов (законов распределения времени обработки запросов) определяются вероятностно-временные характеристики процесса решения задач. Следует отметить, что неотъемлемым свойством непрерывного марковского процесса является экспоненциальность распределения времени пребывания запроса в состоянии s_k ($k = \overline{0, K+1}$). В общем случае законы распределения времени выполнения запросов отличны от экспоненциального. В этом случае процесс решения задач описывается с помощью вложенных марковских цепей (полумарковский процесс). Характеристики полумарковского процесса определяются значительно сложнее, чем для марковского процесса, и вычисление стационарных вероятностей состояний выливается в сложную математическую задачу.

Граф, описывающий процесс решения задачи, может быть интерпретирован разомкнутой сетью массового обслуживания, в которой характеристики СМО (параметры законов распределения времени обработки запросов) определяются на представленной ранее модели. Узел 0 представляет собой источник заявок, а узел k+1 – выход из сети. Конкуренция заявок за ресурсы АИС учитывается во временных параметрах узлов, что позволяет рассматривать процесс решения различных задач, а также одной и той же задачи в различные моменты времени независимо друг от друга.

Как было отмечено, для определения функции распределения случайной величины (в рассматриваемом случае – времени пребывания заявок в сети) достаточно знать первые четыре центральные момента этой величины. Дальнейший переход к искомой функции осуществляется с помощью семейств кривых, охватывающие различные виды распределений, конечных рядов и т.д.

Рассмотрим процесс перемещения заявки по сети. После возникновения в узле 0 заявка с вероятностью p_{01} поступит в узел 1, с вероятностью p_{02} – в узел 2 и т.д., причем $\sum_{j=1}^{h} p_{0j} = 1$, где h – число узлов, смежных с узлом 0, так как. случайные события выбора различных направлений движения заявки из узла составляют полную группу. Аналогично для произвольного узла k $\sum_{j=k+1}^{w} p_{kj} = 1$. Значения вероятностей p_{kl} существенно зависят от интенсивности потока запросов, поступающих в систему. При этом для узла k с альтернативными путями вероятности p_{kl}, p_{km} пропорциональны интенсивностям l-го и m-го потоков запросов.

Для вероятностей выбора путей от узла 0 до узла k+1 также выполняется условие $\sum_{i=1}^{L} p_i = 1$, где p_i – вероятность прохождения заявки по i-му пути, L – число путей.

Из этого следует, что плотность распределения времени пребывания заявки в сети определяется равенством $f(t) = \sum_{i=1}^{L} p_i f_i(t)$, где $f_i(t)$ – плотность распределения времени прохождения заявкой i-го пути [2].

Интегрируя обе части этого равенства в пределах от 0 до $t_{доп}$, и учитывая, что интеграл суммы равен сумме интегралов, получим

$$F(t \le t_{доп}) = \sum_{i=1}^{L} p_i F_i(t \le t_{доп}), \tag{1}$$

где $F_i(t \le t_{доп}) = \int_0^{t_{доп}} f_{\text{i}}(t)dt$.

Имея в виду допущение о независимости времени пребывания заявок на каждом элементе сети, центральные моменты времени прохождения заявкой i-го пути находятся по известным соотношениям [8]:

$$M_{ki}^{(\text{цц}} = \sum_{i=1}^{v_i} m_{ki}^{(\text{ц})}, k = \overline{1,3};$$

$$M_{4i}^{(\text{ц})} = \sum_{j=1}^{v_i} m_{4j}^{(\text{ц})} + 6\sum_{j=1}^{v_i} m_{2j}^{(\text{ц})} \sum_{u=j+1}^{v_i} m_{2u}^{(\text{ц})},$$

где $M_{kj}^{(\text{ц})}$ – центральный момент k-го порядка времени пребывания заявки в j-м узле; v_i – число слагаемых в сумме (узлов, входящих в i-й путь).

Переход от начальных моментов времени пребывания заявок на узле к центральным моментам осуществляется согласно [6]:

$$m_r^{(ц)} = \sum_{s=0}^{r} \binom{r}{s} m_{r-s}^{(н)} \left(-m_1^{(н)}\right)^s,$$

где $\binom{r}{s} = \dfrac{r!}{s!(r-s)!}$.

В частности, для первых четырех моментов:

$m_{1j}^{(ц)} = m_{1j}^{(н)};$	$m_{3j}^{(ц)} = m_{3j}^{(н)} - 3m_{2j}^{(н)}m_{1j}^{(н)} + 2\left(m_{1j}^{(н)}\right)^3;$
$m_{2j}^{(ц)} = m_{2j}^{(н)} - \left(m_{1j}^{(н)}\right)^2;$	$m_{4j}^{(ц)} = m_{4j}^{(н)} - 4m_{3j}^{(н)}m_{1j}^{(н)} + 6m_{2j}^{(н)}\left(m_{1j}^{(н)}\right)^2 - 3\left(m_{1j}^{(н)}\right)^4,$

где $m_{rj}^{(н)}$ – r-й начальный момент случайной величины t – времени пребывания заявки на j-м узле $\left(r = \overline{1,4};\, j = \overline{1,M}\right)$.

Использование взвешенного суммирования (1) требует существенных затрат вычислительных ресурсов для многократного нахождения величины $F_i\left(t \le t_{доп}\right)$. Упростить решение данной задачи позволяют некоторые особенности исследуемой СеМО:

1) в произвольный момент в сети присутствует только одна заявка и никакая другая не поступает на вход до тех пор, пока не будет обслужена предыдущая;

2) при обслуживании заявка продвигается строго по одному пути, определяемому вероятностью его выбора.

Учитывая эти особенности, можно рассматривать сеть как состоящую из независимых путей. Поскольку операция взвешенного суммирования применима по отношению к законам распределения случайных величин, уместно предположить возможность ее использования и для нахождения некоторых других числовых характеристик времени пребывания заявки в сети. Доказательством этой гипотезы выступает следующая теорема.

Теорема. Начальные моменты k-го порядка множества случайных величин равны взвешенной сумме начальных моментов того же порядка входящих в множество величин при условии, что события, заключающиеся в выборе случайных величин, несовместны.

Дано: Множество случайных величин $X_1, X_2, ..., X_m$, для которых указаны их возможные значения $X_j = \{x_i\}_j$ $\left(i = \overline{1,n};\, j = \overline{1,m}\right)$ и соответствующие им вероятности $p_{ij} = P\left(X_j = x_i\right)$.

Требуется доказать: $M_k^{(н)} = \sum_{j=1}^{n} \alpha_j M_{kj}^{(н)}$, где $M_k^{(н)}$ – k-й начальный момент множества случайных величин, $M_{kj}^{(н)}$ – k-й начальный момент j-й случайной величины, входящей в множество, α_j – весовой коэффициент этой величины, выбранный при условии $\sum_{j=1}^{m} \alpha_i = 1$.

Доказательство: По определению начальных моментов $M_{kj}^{(н)} = \sum_{i=1}^{n} x_i^k p_{ij}$. Так как события $\mathrm{X_j} = x_1, X_j = x_2, ..., X_j = x_n$ образуют полную группу попарно несовместных событий, то $\sum_{i=1}^{n} p_{ij} = 1$.

Обобщая на случай множества случайных величин, получим $M_k^{(н)} = \sum_{i=1}^{n} x_i^k p_i$, где $p_i = P[(X_1 \vee X_2 \vee ... \vee X_m) = x_i]$.

Учитывая, что события, заключающиеся в выборе различных случайных величин из множества, несовместны, а выбор одной из величин и выпадение ее возможного значения есть события независимые, можно записать $p_i = \sum_{j=1}^{m} p_j p_{ij}$, где p_j – вероятность выбора случайной величины X_j.

Следовательно,

$$M_k^{(н)} = \sum_{i=1}^{n} x_i^k \sum_{j=1}^{m} p_j p_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i^k p_i p_{ij} = \sum_{j=1}^{m} \sum_{i=1}^{n} x_i^k p_j p_{ij} = \sum_{j=1}^{m} p_j \sum_{i=1}^{n} x_i^k p_{ij} = \sum_{j=1}^{m} p_j M_{kj}^{(н)} .$$

Для несовместных событий выбора случайной величины X_j из множества выполняется условие $\sum_{j=1}^{m} p_j = 1$, поэтому вероятности p_j являются весовыми коэффициентами $\alpha_j \left(j = \overline{1,m}\right)$ и справедливо утверждение $M_k^{(н)} = \sum_{j=1}^{n} p_j M_{kj}^{(н)}$.

Применительно к исследуемой СеМО можно утверждать, что k-й начальный момент времени пребывания заявки в сети равен взвешенной сумме начальных моментов того же порядка времени прохождения каждого из путей, причем в качестве весовых коэффициентов используются вероятности выбора путей.

Значения начальных моментов времени прохождения i-го пути $M_i^{(н)}$ могут быть получены на основе выражения [6]:

$$M_r^{(н)} = \sum_{s=0}^{r} \binom{r}{s} M_{r-s}^{(ц)} \left(M_1^{(н)}\right)^s .$$

Для $r = \overline{1,4}$ имеем

$M_{1i}^{(н)} = M_{1i}^{(ц)}$;	$M_{3i}^{(н)} = M_{3i}^{(ц)} + 3M_{2i}^{(ц)} M_{1i}^{(н)} + \left(M_{1i}^{(н)}\right)^3$;
$M_{2i}^{(н)} = M_{2i}^{(ц)} + \left(M_{1i}^{(н)}\right)^2$;	$M_{4i}^{(н)} = M_{4i}^{(ц)} + 4M_{3i}^{(ц)} M_{1i}^{(н)} + 6M_{2i}^{(ц)} \left(M_{1i}^{(н)}\right)^2 + \left(M_{1i}^{(н)}\right)^4$,

где $M_{ri}^{(н)}$ – r-й начальный момент случайной величины t_i, $M_{ri}^{(ц)}$ – r-й центральный момент той же величины.

Таким образом, дуальность свойств моментов (центральные моменты могут подвергаться сложению, а начальные – взвешенному суммированию) позволяет оперировать только этими числовыми характеристиками при вычислении промежуточных расчетов и однократно вычислять значение функции распределения времени пребывания заявки в СеМО.

Восстановление функции распределения по ее моментам проводится в предположении об унимодальности функции плотности случайной величины [6, 8]. В общем случае суперпозиция унимодальных функций плотности времени прохождения заявкой независимых путей может приводить к наличию многих максимумов. Иначе говоря, результирующая функция плотности является мультимодальной. Тем не менее, уместно предположить, что при большом числе независимых путей суммарная функция плотности сглаживается, стремясь к унимодальной. Показано, что распределения, имеющие конечное число соответственно одинаковых низших моментов (четыре или даже три) в некотором смысле аппроксимируют друг друга [6]. При этом представляется возможным аппроксимировать исследуемое распределение, найдя другое, известного вида с теми же моментами. Принимая полученные моменты за числовые характеристики с унимодальной плотностью, правомочно считать это распределение за искомое.

Восстановление функции распределения конечными рядами часто оказывается неудовлетворительным, особенно в хвостовой части распределения, поэтому более широкое распространение получил метод кривых Пирсона. Подробное описание метода приводится в ряде работ, в том числе в [6, 8]. В основе метода лежит нахождение выражений для функции плотности случайной величины по значениям первых четырех центральных моментов и дальнейшее численное интегрирование этих выражений, так как для основных типов распределений интегралы от функций плотности через элементарные функции не выражаются. Метод не лишен ряда недостатков, связанных с вычислительным аспектом его реализации. К ним относятся:

1) необходимость численного интегрирования в процессе вычислений, что в сущности приводит к возникновению погрешности результата;

2) наличие уникальных выражений для вычисления каждого типа функции распределения, требующее предварительной идентификации типа кривой и существенного расхода памяти ЭВМ.

Попытка устранения перечисленных недостатков явилась причиной поиска других подходов к аппроксимации функции распределения, заключающихся в подборе кривых непосредственно к искомой функции распределения, а не к ее плотности.

В работе [9] показано, что в общем виде функция распределения может быть описана как $F(x) = \{1 + \exp[-G(x)]\}^{-1}$, где $G(x) = \int_0^x g(t)dt$. Здесь $g(t)$ – некоторая функция которая должна быть неотрицательной при $0 \le F(x) \le 1$ и в области изменения *x*. Существует ряд функций, таких как *c*, ct^{-1}, $((c-t)t)$, $c + \sec^2 t$, $c \times ch(t)$, где $c > 0$, для которых интеграл на интервале $-\infty < x < \infty$ таков, что $0 \le F(x) \le 1$. При этом задача аппроксимации сводится к отысканию функции $g(t)$, через которую находятся значения функции $F(x)$. Известны функции, удовлетворяющие признакам функции распределения [9]. Среди них следующие:

$F(x) = x,$	$0 \le x \le 1;$
$F(x) = \left(e^{-x} + 1\right)^{-r},$	$-\infty < x < \infty;$
$F(x) = \left(x^{-k} + 1\right)^{-r},$	$0 \le x < \infty;$
$F(x) = \left(ke^{-tgx} + 1\right)^{-r},$	$\pi/2 \le x \le \pi/2;$
$F(x) = \left(ke^{-cshx} + 1\right)^{-r},$	$-\infty < x < \infty;$
$F(x) = 1 - \left(1 + x^c\right)^{-k},$	$0 \le x < \infty$

и другие, причем *c*, *k* и *r* – некоторые положительные константы.

В общем случае вычисление значений функции распределения включает в себя три этапа:

1) выбор типа функции;
2) определение параметров распределения;
3) расчет значения функции.

Для отыскания функции $F(x)$ требуется найти ее параметры *c*, *k* и *r* через значения числовых характеристик, таких как среднее значение $(\bar{x})$, стандартное отклонение (σ), параметр асимметрии $(\sqrt{\beta_1})$ и параметр эксцесса (β_2). Величины $\sqrt{\beta_1}$ и β_2 находятся по соотношениям:

$$\sqrt{\beta_1} = m_3 / \sqrt{m_2^3}; \qquad \beta_2 = m_4 / m_2^2 .$$

где m_k – *k*-й центральный момент случайной величины $(k = \overline{2,4})$.

В своей сущности метод основан на вычислении кумулятивных моментов, определяемых как $M_j = \int_0^{\infty} x^j [1 - F(x)]dx - \int_{-\infty}^{0} x^j F(x)dx$, и перехода от них к значениям $\bar{x}$, σ, $\sqrt{\beta_1}$ и β_2.

С точки зрения практического использования особый интерес представляет функция

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \left(1 + x^c\right)^{-k}, & x \geq 0 \end{cases}. \qquad (2)$$

Вычисленные для этой функции возможные значения $\sqrt{\beta_1}$ и β_2 образуют широкий диапазон, который охватывает многие распределения. При этом в область этих распределений попадают только функции плотности колоколообразного вида, к которым относятся и функции плотности времени пребывания заявок в СеМО.

Применительно к функции (2) выражение для нахождения кумулятивных моментов выглядит следующим образом:

$$M \ = \frac{\Gamma\left[\dfrac{j+1}{c}\right]\Gamma\left[k - \dfrac{j+1}{c}\right]}{c\Gamma(k)},$$

где Γ – гамма-функция, $j = 0, 1,\ldots$, кроме $j < ck$-1, c и k – действительные числа, лежащие на оси от 1 до ∞.

Полагая $m_j^{(н)} = jM_{j-1} \ (j > 0)$ и $m_j^{(ц)} = j\sum_{j=0}^{j-1}\binom{j-1}{i}(-M_0)^i M_{j-1-i} + (-M_0)^j \ (j > 1)$,

получим соотношения для первых четырех центральных моментов распределения

$m_1^{(ц)} = m_1^{(н)} = M_0$;	$m_3^{(ц)} = 3M_2 - 12M_1M_0 + 2M_0^3$;
$m_2^{(ц)} = 2M_1 - M_0^2$;	$m_4^{(ц)} = 4M_3 - 12M_2M_0 + 12M_1M_0^2 - 3M_0^4$,

от которых осуществляется переход к параметрам $\bar{x}$, σ, $\sqrt{\beta_1}$, β_2 и далее к коэффициентам c и k.

Литература

1. Рыжиков Ю. И., Хомоненко А. Д. Расчет разомкнутых немарковских цепей с преобразованием потоков // АВТ.– 1989.– №3.– С.15–24.
2. Клейнрок Л. Теория массового обслуживания.– М., 1979.– 432 с.
3. Бусленко Н. П. Моделирование сложных систем.– М., 1978.– 399 с.
4. Иванов Е. В. Имитационное моделирование средств и комплексов связи и автоматизации.– СПб., 1992.– 206 с.
5. Демиденко Ю. А., Рыжиков Ю.И. Определение моментов распределения времени пребывания заявки в вычислительной сети. // АВТ.– 1986.– №3.– С.24–27.
6. Кендалл М., Стьюарт А. Теория распределений.– М., 1986.– 588 с.
7. Разработка САПР: В 10 кн. Кн. 8. Математические методы анализа надежности и производительности САПР: Практ. пособие. / В. И. Кузовлев, П. Н. Шкатов / Под ред. А. В. Петрова – М., 1990.– 144 с.
8. Ходасевич Г. Б. Математическое обеспечение АСОД / Под ред. П. И. Волкова. – Л., 1981.– 176 с.
9. Burr I. W. Cumulative Frequency Functions // Ann. Math, Statist., Vol. 13.– Baltimore, 1942. – p.215–232.