УДК 519.217-044.64

Укрупнение состояний марковских процессов на основе частот

Б. П. Зеленцов

Изложен метод укрупнения состояний эргодического однородного марковского процесса в дискретном и непрерывном времени, в основе которого лежат средние частоты переходов между состояниями и подмножествами состояний. Множество состояний исходного процесса разбивается на подмножества, каждое из которых заменяется на одно укрупнённое состояние. Затем находятся характеристики этих подмножеств: средние частоты переходов между подмножествами в стационарном режиме, предельные вероятности подмножеств, средняя продолжительность нахождения в подмножествах состояний, средняя продолжительность цикла каждого подмножества. Эти зависимости находятся путём оперирования с числовыми и функциональными матрицами. Метод может быть применён для моделирования вероятностных систем с большим числом состояний.

Ключевые слова: марковский процесс в дискретном и непрерывном времени, средняя частота переходов между подмножествами состояний, средняя частота подмножества состояний.

1. Введение

Одной из проблем в области моделирования сложных систем является увеличение размерности, которое выражается в увеличении числа состояний системы. Большое число состояний и событий обусловлено разными физическими процессами, разными факторами, среди которых можно отметить фазы эксплуатации (использование по назначению, периодические проверки, восстановление), разные виды неисправностей и отказов, наличие непрерывного и периодического контроля технического состояния и др. Так, в рамках сложных систем исследовано поведение бактерий, совершающих случайные движения, приведены эргодические теоремы и условия эргодичности [18]. В [16] рассмотрена асимптотическая сходимость статистических относительных частот событий, приведены аналитические выражения, описывающие процессы полимеризации и их подтверждение путём статистического моделирования.

В [15] выведены формулы для среднего времени возвращения в подмножество состояний марковского процесса в дискретном и непрерывном времени. Приведены условия существования предельных вероятностей. Переход от процесса в непрерывном времени к вложенной цепи Маркова позволил упростить аналитическую модель и довести её до развёрнутых формул для вероятностей состояний и доказать сходимость вероятностей состояний к предельным [17]. В этих исследованиях увеличение числа состояний системы вызывает естественные трудности на всех этапах реализации модели: при формировании множества состояний системы и исходных характеристик системы приводит к необходимости оперирования с характеристиками большой размерности. В аналитических моделях систем большой размерности возрастает трудоёмкость получения формул, а сами формулы настолько сложны, что их применение в инженерных расчётах практически не оправдано. Зачастую большое число состояний является существенным препятствием при моделировании функционирования систем. В этом случае переходят к алгоритмическому и имитационному моделированию.

В [12, 13] рассмотрена специфическая разложимая структура, названная «почти полностью разложимая цепь». В статье представлены алгоритмы, которые приближённо вычисляют характеристики цепи Маркова в дискретном времени. Алгоритмы основаны на том, что состояния цепи Маркова сгруппированы так, что вероятности изменения состояния внутри группы имеют более высокий порядок по сравнению с вероятностями переходов между группами. Эти алгоритмы позволяют вычислять предельные вероятности состояний, время перехода между состояниями, число попаданий в состояние до поглощения. Предложен древовидный метод вычисления этих характеристик для процесса с большим числом состояний. Метод позволяет приближённо вычислять предельное распределение вероятностей состояний.

Одним из основных путей преодоления трудностей, связанных с большой размерностью сложных систем, является укрупнение состояний, при котором система с данным числом состояний заменяется системой с меньшим числом состояний. Проблема укрупнения – это проблема преобразования первоначальной системы в систему с меньшим числом состояний, изучение которой с точки зрения получения выходных характеристик не приводит к существенной потере информации и искажению результатов моделирования. В связи с этим важное значение имеет разработка новых подходов к укрупнению состояний и развитие известных методов укрупнения состояний, которые не искажают результаты моделирования исходного процесса. При формировании общего подхода к укрупнению состояний следует иметь в виду, что предложения по укрупнению состояний должны быть доказательными, так как произвольное объединение нескольких состояний в одно укрупнённое состояние приводит к потере информации об исходной системе. Существенным обстоятельством является также доведение математически обоснованных методов до практического использования. Проблема укрупнения случайных процессов является актуальной для многих приложений, так как с этой проблемой зачастую связаны возможности моделирования сложных систем.

С точки зрения охвата укрупнением множества состояний укрупнение может быть простым, когда множество состояний разбивается на два подмножества, и блочным, когда множество состояний разбивается на несколько подмножеств (блоков). В [5] приведён частотный метод моделирования случайного процесса при разбиении эргодического множества состояний на два подмножества, при этом вычисляются частоты переходов между этими подмножествами и другие характеристики этих подмножеств. Фактически исходный процесс приведён к двум укрупнённым состояниям. Ключевым понятием здесь является частота переходов между двумя состояниями исходного процесса в стационарном режиме. Средняя частота переходов между подмножествами состояний также является устойчивой характеристикой эргодического марковского процесса в стационарном режиме. Такой вид простого укрупнения можно было бы назвать «частотным укрупнением» состояний марковского процесса в дискретном и непрерывном времени. Этот подход может быть использован для моделирования систем длительного использования.

Необходимое и достаточное условие укрупнения для марковского процесса в дискретном и непрерывном времени сформулировано в ряде работ [1, 7, 8, 9]. При разбиении множества состояний на подмножества исходная матрица переходных вероятностей или интенсивностей представляется в виде соответствующей блочной матрицы. Если сумма элементов каждой строки блока является постоянной, то есть не зависит от номера строки этого блока, то такое множество состояний является укрупнимым: подмножество может быть заменено на одно состояние. Сумма элементов строки блока является переходной вероятностью или интенсивностью перехода нового, укрупнённого, процесса. Такой подход позволяет «склеивать» состояния подмножеств и сохранить точные значения предельных вероятностей укрупнённой системы, при этом предельная вероятность нового, укрупнённого, состояния получается суммированием предельных вероятностей состояний соответствующих укрупняемых состояний исходного процесса. Однако практическое применение сформулированного необходимого и достаточного условия укрупнимости затрудняется тем, что эта идея не доведена до метода, пригодного для применения в инженерной практике.

При большом числе состояний разбиение исходного множества на два подмножества не всегда является эффективным, оно не учитывает особенности и различия между состояниями одного подмножества. Матричные характеристики подмножеств при этом могут оказаться также больших размеров. Поэтому разбиение множества состояний на несколько подмножеств (блочное укрупнение) может найти применение при моделировании сложных систем.

В [14] рассмотрен процесс Маркова в дискретном и непрерывном времени, множество состояний которого разбито на укрупнённые состояния, каждому из которых приписана предельная вероятность. Достаточное условие укрупнимости представлено как разновидность слабой укрупнимости, при которой вероятности исходного процесса совпадают с вероятностями укрупнённого процесса при определённых начальных условиях. Исследована зависимость свойств укрупнённого процесса от начальных условий процесса. Показано, что расширение процесса Маркова также зависит от начальных условий процесса. Разработанная модель применима для биохимических реакций.

В статье [2] представлены алгоритмы проверки возможности укрупнения цепи Маркова и получения закона формирования укрупнённой цепи. Предложена оценка количества возможных пространств, получаемых в процессе укрупнения состояний исходной цепи.

В [4] изложен метод укрупнения состояний системы, эволюционирование которой описывается марковским процессом как в дискретном, так и в непрерывном времени. В соответствии с этим методом система с данным множеством состояний приводится к множеству с меньшим числом состояний путём введения граничных состояний подмножеств. Переходы между состояниями на исходном множестве заменены на переходы между граничными состояниями. Модель позволяет оперировать с матрицами меньших размеров. Информация об исходном множестве состояний определённым образом упакована в систему с меньшим числом состояний с сохранением средней продолжительности нахождения в подмножестве состояний.

Существенным фактором при укрупнении состояний является возможность сохранения точных значений характеристик процесса. В связи с этим предложено относить укрупнение состояний к одному из двух типов: точному (эквивалентному) или приближённому (квазиэквивалентному) [11]. В этой работе исследованы условия, при которых приближённые значении вероятностей укрупнённых состояний стремятся к точным значениям. К сожалению, оценка погрешности при квазиэквивалентном укрупнении не произведена.

Целью настоящей работы является разработка математической модели для укрупнённого процесса на основе использования средних частот переходов между состояниями этого процесса в стационарном режиме, который протекает в дискретном или непрерывном времени. Множество состояний исходного процесса разбивается на несколько подмножеств. По механизму уменьшения числа состояний здесь производится «склеивание» состояний, при котором несколько состояний каждого подмножества заменяются на одно, при этом сохраняется предельная вероятность и средняя частота этого подмножества. Информация об исходном множестве состояний определенным образом упаковывается в процесс с меньшим числом состояний с сохранением определенных выходных характеристик процесса. Такое укрупнение состояний по терминологии [11] можно назвать эквивалентным.

2. Частоты переходов между состояниями исходного процесса

Рассматривается случайный процесс на дискретном эргодическом множестве с конечным числом состояний $S = \{s_i\}$ в дискретном и непрерывном времени. Исходными характеристиками для процесса в дискретном времени являются переходные вероятности p_{ij} между двумя состояниями s_i и s_j на одном шаге, а для процесса в непрерывном времени – интенсивности переходов λ_{ij} между двумя состояниями s_i и s_j. Для этих процессов вводится понятие частоты ω_{ij} непосредственных переходов между двумя состояниями исходного процесса. Для процесса в дискретном времени ω_{ij} – это среднее число непосредственных переходов $s_i \to s_j$, приходящееся на один шаг, а для процесса в непрерывном времени ω_{ij} – это среднее число переходов $s_i \to s_j$, приходящееся на единицу времени.

Частота непосредственных переходов $s_i \to s_j$ следующим образом связана с предельными вероятностями состояний и исходными характеристиками для процесса в дискретном и непрерывном времени [3]:

$$\text{а) } \omega_{ij} = \pi_i \cdot p_{ij}\text{; б) } \omega_{ij} = \pi_i \cdot \lambda_{ij},\ i \neq j, \tag{1}$$

где π_i – предельная вероятность состояния s_i.

Зависимость между параметром ω_{ii} и предельной вероятностью состояния для процесса в дискретном и непрерывном времени определена следующим образом [3]:

$$\text{а) } \omega_{ii} = \pi_i \cdot (p_{ii} - 1)\text{; б) } \omega_{ii} = \pi_i \cdot \lambda_{ii}, \tag{2}$$

где p_{ii} – вероятность остаться в состоянии s_i на одном шаге после попадания в него; λ_{ii} – суммарная интенсивность выхода из состояния s_i, взятая со знаком минус:

$$\text{а) } 1 - p_{ii} = \sum_{j,\, j\neq i} p_{ij}; \quad \text{б) } \lambda_{ii} = -\sum_{j,\, j\neq i} \lambda_{ij}. \tag{3}$$

Формулы для вычисления предельных вероятностей эргодического процесса в дискретном и непрерывном времени приведены в [3].

Будем записывать исходные характеристики процесса в виде матрицы переходных вероятностей $P = \|p_{ij}\|$ для процесса в дискретном времени и матрицы интенсивностей $\Lambda = \|\lambda_{ij}\|$ для процесса в непрерывном времени, а частоты переходов – в виде матрицы частот $\Omega = \|\omega_{ij}\|$. Приведённые соотношения (1) и (2) позволяют выразить зависимость между частотами, предельными вероятностями и исходными характеристиками случайного процесса в дискретном и непрерывном времени в матричном виде [5]:

$$\text{а) } \Omega = \Pi_{\text{dg}} \cdot (P - E)\text{; б) } \Omega = \Pi_{\text{dg}} \cdot \Lambda, \tag{4}$$

где Ω – матрица частот исходного процесса; Π_{dg} – диагональная матрица предельных вероятностей состояний; E – единичная матрица соответствующего порядка.

Итак, соотношения (4) связывают три матричные характеристики исходного множества состояний: частоты переходов между состояниями, предельные вероятности состояний и исходные характеристики случайного процесса (переходные вероятности и интенсивности переходов).

3. Частоты состояний исходного процесса

Введём понятия «частота выхода из состояния s_i», которая представляет собой сумму частот всех переходов из этого состояния в другие состояния, и «частота входа в состояние s_i», которая является суммой частот всех переходов в состояние w_i из других состояний:

$$\omega_i^- = \sum_{j,\, j\neq i} \omega_{ij}; \quad \omega_i^+ = \sum_{j,\, j\neq i} \omega_{ji}, \tag{5}$$

где ω_i^- – частота выхода из состояния s_i, ω_i^+ – частота входа в состояние s_i.

В соответствии с (2) и (3) частота выхода из состояния s_i связана с параметром ω_{ii} следующим образом:

$$\omega_i^- = \sum_{j,\, j\neq i} \omega_{ij} = -\omega_{ii}. \tag{6}$$

Покажем, что $\omega_i^- = \omega_i^+$.

Для процесса в дискретном времени из зависимости для эргодического случайного процесса $\bar{\pi} \cdot P = \bar{\pi}$ [3] следует

$$\bar{\pi} \cdot (P - E) = \bar{o}, \qquad (7)$$

где $\bar{\pi}$ – строка предельных вероятностей исходного процесса; $\bar{o}$ – строка, все элементы которой равны нулю. Выделим на множестве S состояние s_i и выполним умножение (7). В результате получим:

$$\pi_i \cdot (p_{ii} - 1) + \bar{\pi}_i \cdot \dot{p}_i = 0, \qquad (8)$$

где $\bar{\pi}_i$ – строка предельных вероятностей состояний множества S без элемента π_i; $\dot{p}_i$ – i-й столбец матрицы P без диагонального элемента p_{ii}. Из последнего равенства следует:

$$\bar{\pi}_i \cdot \dot{p}_i = -\pi_i \cdot (p_{ii} - 1). \qquad (9)$$

Левая часть последнего равенства представляет собой частоту входа в состояние s_i:

$$\bar{\pi}_i \cdot \dot{p}_i = \sum_{j,\, j \neq i} \pi_j \cdot p_{ji} = \sum_{j,\, j \neq i} \omega_{ji} = \omega_i^+ = -\omega_{ii}. \qquad (10)$$

Сравнивая (6) и (10), видим, что $\omega_i^- = \omega_i^+$, так как $\omega_i^- = -\omega_{ii}$ и $\omega_i^+ = -\omega_{ii}$.

Для процесса в непрерывном времени из соотношения $\bar{\pi} \cdot \Lambda = \bar{o}$ [3] следует:

$$\pi_i \cdot \lambda_{ii} + \bar{\pi}_i \cdot \dot{\lambda}_i = 0, \qquad (11)$$

где $\dot{\lambda}_i$ – i-й столбец матрицы Λ без элемента λ_{ii}. Из этого равенства следует:

$$\bar{\pi}_i \cdot \dot{\lambda}_i = -\pi_i \cdot \lambda_{ii}. \qquad (12)$$

Левая часть последнего равенства является частотой входа в состояние s_i:

$$\bar{\pi}_i \cdot \dot{\lambda}_i = \sum_{j,\, j \neq i} \pi_j \cdot \lambda_{ji} = \sum_{j,\, j \neq i} \omega_{ji} = \omega_i^+. \qquad (13)$$

Сравнив (6) и (13), получим: $\omega_i^- = \omega_i^+$.

Итак, доказано, что $\omega_i^- = \omega_i^+$ для процесса в дискретном и непрерывном времени. Это равенство показывает баланс частот, который заключается в том, что средняя частота вхождения в любое состояние эргодического процесса в стационарном режиме равна средней частоте выхода из этого состояния.

Будем называть среднюю частоту $\omega_i = \omega_i^- = \omega_i^+$ частотой состояния s_i. Очевидно, что $\omega_i = -\omega_{ii}$. Частоты переходов ω_{ij} и частоты состояний ω_i являются устойчивыми характеристиками случайного процесса в стационарном режиме. При изменении этих частот во времени процесс становится нестационарным.

4. Выражение частот переходов через другие характеристики

Из (1) следует, что частота переходов ω_{ij}, $i \neq j$, пропорциональна исходным характеристикам процессов: переходным вероятностям p_{ij} и интенсивностям λ_{ij}. Поскольку частота состояния является суммой частот выхода или суммой частот входа в состояние, то каждая из этих составляющих пропорциональна исходной характеристике перехода. Поэтому частоту перехода можно выразить через частоту состояния и вероятность прохождения $\hat{p}_{ij}$ [3], которая пропорциональна соответствующей исходной характеристике перехода для процесса в дискретном и непрерывном времени:

$$\text{а) } \omega_{ij} = \omega_i \cdot \hat{p}_{ij} = \omega_i \cdot p_{ij} / (1 - p_{ii})\,; \quad \text{б) } \omega_{ij} = \omega_i \cdot \hat{p}_{ij}, = -\omega_i \cdot \lambda_{ij} / \lambda_{ii}, \quad i \neq j, \qquad (14)$$

где $\hat{p}_{ij}$ – вероятности прохождения. Вероятность прохождения $\hat{p}_{ij}$ – это вероятность непосредственного перехода $s_i \to s_j$ при условии, что происходит выход из состояния s_i.

Частоту перехода ω_{ij} можно выразить также через среднюю продолжительность нахождения в состоянии s_i после попадания в него:

$$\text{а) } \omega_{ij} = \omega_i \cdot n_i \cdot p_{ij};; \quad \text{б) } \omega_{ij} = -\omega_i \cdot t_i \cdot \lambda_{ij}, \ i \neq j, \tag{15}$$

где $n_i = 1/(1 - p_{ii})$ – среднее число шагов нахождения в состоянии s_i после попадания в него; $t_i = -1/\lambda_{ii}$ – среднее время нахождения в состоянии s_i после попадания в него.

Представляет интерес выражение частот переходов и частот состояний через исходные характеристики. Подставив в (1) выражения для предельных вероятностей, получим формулы для этих частот. Вычисления могут быть произведены по формулам двух типов.

Первый тип формул основан на вычислении определителей. Частоты переходов для процесса в дискретном и непрерывном времени:

$$\text{а) } \omega_{ij} = \Delta_i \cdot p_{ij} / \Delta;; \quad \text{б) } \omega_{ij} = \Delta_i \cdot \lambda_{ij} / \Delta; \ i \neq j, \tag{16}$$

и частоты состояний для процесса в дискретном и непрерывном времени:

$$\text{а) } \omega_i = \Delta_i \cdot (1 - p_{ij}) / \Delta; \quad \text{б) } \omega_i = -\Delta_i \cdot \lambda_{ii} / \Delta, \tag{17}$$

где $\Delta_i = |E - P_i|$ для процесса в дискретном времени; E – единичная матрица соответствующего порядка; $\Delta_i = |\Lambda_i|$ для процесса в непрерывном времени; P_i – матрица, полученная из матрицы P удалением из неё i-й строки и i-го столбца; Λ_i – матрица, полученная из матрицы Λ удалением из неё i-й строки и i-го столбца; $\Delta = \sum_{i=1}^{m} \Delta_i$ сумма определителей по всем состояниям; m – число состояний исходного процесса.

Второй тип формул основан на обращении матриц. Эти формулы для процесса в дискретном и непрерывном времени:

$$\text{а) } \omega_{ij} = p_{ij} / [1 + \bar{p}_i \cdot (E - P_i)^{-1} \cdot \dot{e}]; \ \text{б) } \omega_{ij} = \lambda_{ij} / [1 - \bar{\lambda}_i \cdot \Lambda_i^{-1} \cdot \dot{e}]; \ i \neq j, \tag{18}$$

где $\bar{p}_i$ – i-я строка матрицы P без элемента p_{ii}; $\bar{\lambda}_i$ – i-я строка матрицы Λ без элемента λ_{ii}; $\dot{e}$ – столбец, все элементы которого равны 1.

5. Матрица частот исходного множества и её свойства

Полученные соотношения для частот переходов между состояниями и частот состояний позволяют выразить их в матричном виде для процесса в дискретном и непрерывном времени:

$$\text{а) } \Omega = D \cdot (P - E) / (\bar{e} \cdot D \cdot \dot{e}); \ \text{б) } \Omega = D \cdot \Lambda / (\bar{e} \cdot D \cdot \dot{e}), \tag{19}$$

где $D = \|d_{ij}\|$ – диагональная матрица, диагональные элементы которой равны соответствующим определителям: $d_{ij} = 0$ для $i \neq j$ и $d_{ii} = \Delta_i$; $\bar{e}$ – строка, все элементы которой равны 1.

Матрица частот может быть вычислена также на основе обращения матриц:

$$\text{а) } \Omega = R \cdot (P - E); \ \text{б) } \Omega = R \cdot \Lambda, \tag{20}$$

где $R = \|r_{ij}\|$ – диагональная матрица, то есть $r_{ij} = 0$ для $i \neq j$. Диагональные элементы матрицы R для процесса в дискретном и непрерывном времени:

$$\text{а) } r_{ii} = [1 + \bar{p}_i \cdot (E - P_i)^{-1} \cdot \dot{e}]^{-1}; \quad \text{б) } r_{ii} = [1 - \bar{\lambda}_i \cdot \Lambda_i^{-1} \cdot \dot{e}]^{-1}. \tag{21}$$

Из определения диагональных элементов матрицы частот Ω и из баланса частот относительно каждого состояния можно сформулировать свойства матрицы частот для процесса в дискретном и непрерывном времени. Поскольку $\omega_{ii} = -\omega_i^-$, то сумма элементов любой строки матрицы Ω равна нулю. Аналогично: поскольку $\omega_{ii} = -\omega_i^+$, то сумма элементов любого столбца матрицы Ω равна нулю. Из этих понятий вытекают свойства матрицы частот для процесса в дискретном и непрерывном времени:

$$\text{а) } \Omega \cdot \dot{e} = \dot{o}; \quad \text{б) } \bar{e} \cdot \Omega = \bar{o}; \quad \text{в) } \bar{e} \cdot \Omega \cdot \dot{e} = 0; \quad \text{г) } |\Omega| = 0, \tag{22}$$

где $\dot{o}$ – столбец, все элементы которого равны 0; $\bar{e}$ – строка, все элементы которой равны 1; $\bar{o}$ – строка, все элементы которой равны 0. Отметим, что столбцы $\dot{e}$ суммируют частоты выхода из состояний, а строки $\bar{e}$ суммируют частоты входа в состояния. Из свойств а) и б) вытекают свойства в) и г): сумма всех элементов матрицы частот равна нулю и определитель этой матрицы равен нулю.

Свойства (22) можно использовать при различных преобразованиях исходного процесса, а также для контроля правильности вычисления матрицы частот.

6. Частотные характеристики подмножеств состояний

Пусть множество состояний S разбито на подмножества по некоторому признаку: $S = \{SI\}$, $I = 1, 2,\ldots, M$, где M – число подмножеств.

Следует отметить, что при использовании данного метода для моделирования систем конкретного назначения разбиению множества S на подмножества целесообразно придавать содержательный смысл. Например, при расчёте надёжности оборудования систем коммутации можно проводить объединение состояний по функциональному признаку, а именно, объединять состояния, если эффект от функционирования системы в одном из них не отличается от эффекта от функционирования в другом с точки зрения решаемой задачи. Например, в [10] в одно подмножество объединены состояния, являющиеся следствием трех видов отказов. Приведённая модель позволила получить частоты этих отказов.

Представим исходные матрицы, а также диагональную матрицу предельных вероятностей и матрицу частот в виде соответствующих блочных матриц: $P = \|P_{IJ}\|$; $\Lambda = \|\Lambda_{IJ}\|$; $\Pi_{dg} = \|\Pi_{dgI}\|$; $\Omega = \|\Omega_{IJ}\|$. Подматрицы P_{IJ}, Λ_{IJ}, Ω_{IJ} имеют размеры $m_I \times m_J$, где m_I – число состояний подмножества SI, m_J – число состояний подмножества SJ. Выполнив умножение (4) в виде блочных матриц, получим для процесса в дискретном и непрерывном времени:

$$\left.\begin{array}{l} \text{а) } \Omega_{IJ} = \Pi_{dgI} \cdot P_{IJ}; \quad \text{б) } \Omega_{IJ} = \Pi_{dgI} \cdot \Lambda_{IJ}, \quad I, J = 1, 2, \ldots, M, \quad I \neq J \\ \text{а) } \Omega_{II} = \Pi_{dgI} \cdot (P_{II} - E); \quad \text{б) } \Omega_{II} = \Pi_{dgI} \cdot \Lambda_{II}, \; I = J, \; I = 1, 2, \ldots, M \end{array}\right\}. \tag{23}$$

Найдём частоты переходов между состояниями подмножеств SI и SJ. Умножим подматрицу частот Ω_{IJ} сначала на строку $\bar{e}$ и затем на столбец $\dot{e}$, которые имеют согласованные размеры с подматрицей Ω_{IJ}:

$$\bar{\omega}IJ = \bar{e} \cdot \Omega_{IJ}; \; \dot{\omega}IJ = \Omega_{IJ} \cdot \dot{e}, \quad I \neq J, \tag{24}$$

где $\bar{e}$ – строка, все элементы которой равны 1; $\dot{e}$ – столбец, все элементы которого равны 1. Строка $\bar{\omega}IJ$ состоит из m_J элементов: каждый элемент является частотой переходов из подмножества SI в состояния подмножества SJ. Столбец $\dot{\omega}IJ$ состоит из m_I элементов: каждый элемент является частотой переходов из состояний подмножества SI в подмножество SJ.

Частоты переходов между состояниями исходного процесса являются безусловными характеристиками переходов. Поэтому частота переходов из подмножества SI в подмножество SJ определяется суммированием всех элементов строки $\bar{\omega}_{IJ}$, или суммированием всех элементов столбца $\dot{\omega}_{IJ}$, или суммированием всех элементов подматрицы Ω_{IJ}:

$$\omega IJ = \bar{\omega}IJ \cdot \dot{e} = \bar{e} \cdot \dot{\omega}IJ = \bar{e} \cdot \Omega_{IJ} \cdot \dot{e}, \; I \neq J, \tag{25}$$

где ωIJ – частота непосредственных переходов $SI \rightarrow SJ$, которая имеет такое же определение, как частота ω_{ij} для исходных процессов в дискретном и непрерывном времени.

Таким образом, частоты переходов являются безусловными характеристиками переходов как между состояниями исходного процесса, так и между подмножествами состояний укрупнённого процесса. Это обстоятельство позволяет получать частоты переходов между подмножествами состояний путём суммирования частот переходов между соответствующими состояниями исходного процесса.

По аналогии с частотами состояний исходного множества введём понятия «частота выхода из подмножества» и «частота вхождения в подмножество состояний»:

$$\omega I^- = \sum_{J,J\neq I} \omega IJ; \quad \omega I^+ = \sum_{J,J\neq I} \omega JI \ , \tag{26}$$

где ωI^- – частота выхода из подмножества W_I, ωI^+ – частота вхождения в подмножество W_I.

Частота ωI^- является частотой переходов из подмножества SI в другие подмножества и определяется путём суммирования частот переходов из подмножества SI в другие подмножества. Частота ωI^+ – это частота переходов из других подмножеств в подмножество W_I, она равна сумме частот переходов из других подмножеств в подмножество SI. В [5] показано, что при разбиении множества состояний на два подмножества средняя частота входа в подмножество состояний равна средней частоте выхода из него, то есть $\omega I^+ = \omega I^-$. После каждого вхождения в подмножество SI следует выход из него и каждому выходу из подмножества SI предшествует вхождение в него. Поэтому частота вхождений в подмножество состояний равна частоте выхода из него. В этом заключается баланс частот для укрупнённого множества состояний.

Будем называть частоту $\omega I = \omega I^+ = \omega I^-$ частотой подмножества SI.

7. Матрица частот укрупнённого процесса

Укрупнённый процесс – это процесс, в котором каждое подмножество SI заменено на одно состояние. Будем обозначать характеристики укрупнённого процесса теми же символами, что и исходный процесс, а номера состояний обозначены строчными буквами.

Представим частоты ωIJ в виде матрицы частот переходов между подмножествами состояний $\Omega_{\text{укр}} = \|\omega IJ\|$, где $\omega IJ = -\omega I$. Поскольку частоты ωIJ на укрупнённом множестве состояний образованы путём суммирования элементов подматриц Ω_{IJ}, то матрица частот укрупнённого множества обладает теми же свойствами, что и матрица частот исходного множества. Эти свойства определяются балансом частот для процессов в стационарном режиме. Таким образом, для укрупнённого процесса

$$\text{а) } \Omega_{\text{укр}} \cdot \dot{e} = \dot{o}; \ \text{б) } \bar{e} \cdot \mathbf{\Omega} = \bar{o}; \ \text{в) } \bar{e} \cdot \Omega_{\text{укр}} \cdot \dot{e} = 0; \ \text{г) } \left|\Omega_{\text{укр}}\right| = 0 . \tag{27}$$

Замечание. Баланс частот для укрупнённого процесса может быть доказан так же, как и для исходного процесса.

8. Циклы подмножеств состояний укрупнённого процесса

Рассмотрим циклическое функционирование случайного процесса относительно каждого подмножества. Смысл такого функционирования заключается в том, что в случайные моменты времени имеют место вхождения в каждое подмножество SI и выходы из него.

Разобьём множество состояний S на два подмножества: SI и $\overline{SI}$, где $\overline{SI}$ – дополнение подмножества SI до множества S. Введём понятие SI-цикла, который заключается в пребывании процесса в подмножестве SI и следующем за ним пребывании в подмножестве $\overline{SI}$. Очевидно, что каждое подмножество SI имеет свой «индивидуальный» SI-цикл. Таким образом, время протекания процесса в стационарном режиме разбивается на циклы относительно каждого подмножества, состоящие из пребывания процесса в подмножестве SI и следующего за ним пребывания в других подмножествах: $SI \to \bar{S}I \to SI \to \bar{S}I \to \ldots$

Введём обозначения: nI и $n\bar{I}$ – среднее число шагов нахождения в подмножествах SI и $\bar{S}I$ на SI-цикле для процесса в дискретном времени; tI и $t\bar{I}$ – среднее время нахождения в подмножествах SI и $\bar{S}I$ на SI-цикле для процесса в непрерывном времени. Тогда средняя продолжительность SI-цикла для процесса в дискретном и непрерывном времени будет иметь вид:

$$\text{а) } n_{\text{ц}}I = nI + n\bar{I}\,; \quad \text{б) } t_{\text{ц}}I = tI + t\bar{I}\,. \qquad (28)$$

Замечание. Средняя продолжительность SI-цикла – это средний период между вхождениями в подмножество SI.

Очевидно, что частота подмножества SI (ωI) является частотой SI-цикла. Эта частота может быть вычислена как по частоте выхода из подмножества SI, так и по частоте входа в это подмножество. Итак, частота цикла ωI и средняя продолжительность SI-цикла для процесса в дискретном и непрерывном времени связаны следующим образом:

$$\text{а) } n_{\text{ц}}I = 1/\omega I\,; \quad \text{б) } t_{\text{ц}}I = 1/\omega I\,,\ I = 1, 2, \ldots, M. \qquad (29)$$

Предельные вероятности состояний являются безусловными суммируемыми характеристиками случайного процесса. Поэтому предельная вероятность подмножества SI определяется суммой предельных вероятностей состояний этого подмножества как для процесса в дискретном, так и в непрерывном времени:

$$\pi I = \bar{e} \cdot \Pi_{\text{dg}I} \cdot \dot{e}, \qquad (30)$$

где πI – предельная вероятность подмножества SI, имеющая смысл средней доли продолжительности нахождения в подмножестве SI среди всех подмножеств при бесконечном протекании процесса.

В основе подхода к вычислению характеристик укрупнённого процесса лежит свойство суммируемости предельных вероятностей состояний исходного процесса и частот переходов между состояниями исходного процесса, то есть частота переходов между подмножествами SI и SJ определяется по (25), а предельная вероятность подмножества SI определяется по (30).

Средняя продолжительность нахождения в подмножестве SI после попадания в него определяется как часть SI-цикла с вероятностью πI для процесса в дискретном и непрерывном времени:

$$\text{а) } nI = \pi I \cdot n_{\text{ц}}I = \pi I / \omega I; \quad \text{б) } tI = \pi I \cdot t_{\text{ц}}I = \pi I / \omega I. \qquad (31)$$

Можно провести рассуждения, аналогичные для исходного процесса. За достаточно большой интервал времени [0; t] в стационарном режиме среднее время нахождения в подмножестве SI для процесса в непрерывном времени составит $\pi I \cdot t$. С другой стороны, среднее число вхождений в подмножество SI составит $\omega I \cdot t$. При каждом вхождении в подмножество SI среднее время пребывания в нём составит tI. За время t средняя продолжительность нахождения в подмножестве SI составит $\omega I \cdot t \cdot tI$. Приравняв средние времена, получим: $\pi I = \omega I \cdot tI$.

Аналогичные рассуждения можно провести для процесса в дискретном времени.

Итак, можно записать зависимость между предельной вероятностью подмножества, частотой подмножества и средней продолжительностью нахождения в подмножестве состояний для процесса в дискретном и непрерывном времени:

$$\text{а) } \pi I = \omega I \cdot nI. \quad \text{б) } \pi I = \omega I \cdot tI. \qquad (32)$$

Замечание. В [3] средняя продолжительность нахождения в подмножестве состояний найдена с учетом начального распределения вероятностей состояний при вхождении в подмножество.

9. Частный случай выделения одного состояния

Пусть подмножество SI состоит из одного состояния $SI = \{s_i\}$. Тогда $\bar{S}_i$ – множество возможных состояний без состояния s_i. В этом случае s_i-цикл заключается в пребывании процесса в состоянии s_i и следующем за ним пребывании в подмножестве $\bar{S}_i$. Из этого следует, что частота s_i-цикла является частотой состояния s_i. Характеристики состояния s_i и соответствующего s_i-цикла приведены в табл. 1.

Очевидно, что для одного выделенного состояния сохраняются характеристики как в исходном, так и в укрупнённом варианте: π_i, p_{ii}, λ_{ii}.

Таблица 1. Характеристики состояния s_i и соответствующего s_i-цикла

Характеристика	Процесс	
	в дискретном времени	в непрерывном времени
1. Предельная вероятность состояния s_i	π_i	π_i
2. Исходная характеристика состояния s_i	p_{ii}	λ_{ii}
3. Частота состояния s_i	$\omega_{ii} = \pi_i \cdot (1 - p_{ii})$	$\omega_{ii} = -\pi_i \cdot \lambda_{ii}$
4. Средняя продолжительность s_i-цикла	$n_{\text{ц}}I = \dfrac{1}{\pi_i \cdot (1 - p_{ii})}$	$t_{\text{ц}}I = -\dfrac{1}{\pi_i \cdot \lambda_{ii}}$
5. Средняя продолжительность состояния s_i	$n_i = \dfrac{1}{1 - p_{ii}}$	$t_i = -\dfrac{1}{\lambda_{ii}}$
6. Средняя продолжительность подмножества $\bar{S}_i$	$n_{\bar{i}} = \dfrac{1 - \pi_i}{\pi_i \cdot (1 - p_{ii})}$	$t_{\bar{i}} = -\dfrac{1 - \pi_i}{\pi_i \cdot \lambda_{ii}}$

10. Показатели укрупнимости

Эффект от уменьшения числа состояний можно оценить коэффициентом укрупнимости, который представляет собой число «исчезнувших», или «упакованных», состояний при замене исходного процесса на укрупнённый, приходящееся на одно укрупнённое состояние:

$$L = (m - M)/M , \tag{33}$$

где m – число состояний исходного процесса, M – число состояний укрупненного процесса.

Число «исчезнувших» состояний – это разность между числом состояний исходного процесса и числом состояний укрупнённого процесса. Чем больше эта разность, тем выше эффект от укрупнения. При приведении числа «исчезнувших» состояний к одному укрупнённому состоянию можно использовать коэффициент укрупнимости для сравнения эффекта от укрупнения для разных вариантов укрупнения одного процесса и для сравнения эффекта от укрупнения для разных случайных процессов.

Коэффициент укрупнимости является удобным показателем для практического применения, поскольку он характеризует эффект укрупнения, а именно, коэффициент укрупнимости возрастает с увеличением числа «упакованных» состояний. Следует иметь в виду, что при отсутствии укрупнения коэффициент укрупнимости равен нулю.

11. Пример укрупнения

Приведем простой иллюстративный пример с небольшим числом состояний на основе модели функционирования линии связи в условиях недостоверного контроля технического состояния [6]. В модель заложены пять пронумерованных состояний:

1 – работоспособное состояние линии связи;

2 – неработоспособное состояние линии связи;

3 – проверка линии связи, находящейся в работоспособном состоянии;

4 – проверка линии связи, находящейся в неработоспособном состоянии;

5 – восстановление линии связи.

Процесс протекает в непрерывном времени, матрица интенсивностей имеет вид:

$$\Lambda = \begin{pmatrix} -\lambda-\gamma & \lambda & \gamma & 0 & 0 \\ 0 & -\gamma & 0 & \gamma & 0 \\ (1-\alpha)\cdot\mu_{\text{п}} & 0 & -\mu_{\text{п}} & 0 & \alpha\cdot\mu_{\text{п}} \\ 0 & \beta\cdot\mu_{\text{п}} & 0 & -\mu_{\text{п}} & (1-\beta)\cdot\mu_{\text{п}} \\ \mu_{\text{в}} & 0 & 0 & 0 & -\mu_{\text{в}} \end{pmatrix}.$$

В матрице обозначено: λ – интенсивность отказов линии связи; γ – интенсивность начала периодической проверки со случайным периодом; μ – интенсивность завершения периодической проверки; α – вероятность ошибки контроля I рода; β – вероятность ошибки контроля II рода; $\mu_{\text{в}}$ – интенсивность завершения восстановления.

Граф состояний может быть легко восстановлен по приведённой матрице интенсивностей.

Пример приведён в числах, поскольку формулы являются громоздкими. Исходные данные взяты такие же, как в [6], а именно: средняя наработка на отказ – 10 000 часов, среднее время между проверками – 72 секунды, среднее время восстановления – 2 часа, средняя продолжительность одной проверки – 1 с. Исходя из этих значений приняты следующие значения интенсивностей событий, приведённых к единице измерения час$^{-1}$:

интенсивность отказов: $\lambda = 10^{-4}$;

интенсивность начала проверки: $\gamma = 50$;

интенсивность завершения проверки: $\mu_{\text{п}} = 3\,600$;

интенсивность завершения восстановления: $\mu_{\text{в}} = 0.5$;

вероятность ошибки контроля I рода: $\alpha = 0.001$;

вероятность ошибки контроля II рода: $\beta = 0.001$.

Строка предельных вероятностей, вычисленных по формулам на основе определителя или на основе обращения матриц [3]:

$$\bar{\pi} = \begin{pmatrix} 0.898 & 1.797\cdot 10^{-6} & 0.012 & 2.496\cdot 10^{-8} & 0.09 \end{pmatrix}.$$

Матрица частот, вычисленная по (19) или (20):

$$\Omega = \begin{bmatrix} -44.88 & 8.976\cdot 10^{-5} & 44.88 & 0 & 0 \\ 0 & -8.985\cdot 10^{-5} & 0 & 8.985\cdot 10^{-5} & 0 \\ 44.835 & 0 & -44.88 & 0 & 0.045 \\ 0 & 8.985\cdot 10^{-8} & 0 & -8.985\cdot 10^{-5} & 8.976\cdot 10^{-5} \\ 0.045 & 0 & 0 & 0 & -0.045 \end{bmatrix}.$$

Свойства (22) матрицы частот выполняются (в данном примере это можно проверить также и визуально).

Примечание. При вычислениях с помощью компьютерных средств (Mathcad, MATLAB) при проверке свойств матрицы частот зачастую получают не нули, а числа, близкие к нулю и имеющие порядок 10^{-14} или 10^{-15}.

Разбиение множества состояний по функциональному признаку на два подмножества: $S1 = \{1, 2, 3, 4\}$; $S2 = \{5\}$. В соответствии с этим матрица частот разбивается на четыре подматрицы:

$$\Omega_{11} = \begin{pmatrix} -44.88 & 8.976\cdot10^{-5} & 44.88 & 0 \\ 0 & -8.985\cdot10^{-5} & 0 & 8.985\cdot10^{-5} \\ 44.835 & 0 & -44.88 & 0 \\ 0 & 8.985\cdot10^{-8} & 0 & -8.985\cdot10^{-5} \end{pmatrix}; \quad \Omega_{12} = \begin{pmatrix} 0 \\ 0 \\ 0.045 \\ 8.976\cdot10^{-5} \end{pmatrix};$$

$$\Omega_{21} = \begin{pmatrix} 0.045 & 0 & 0 & 0 \end{pmatrix}; \qquad \Omega_{22} = \left(-0.045\right).$$

Частоты переходов между подмножествами $S1$ и $S2$, вычисленные по (25):

$$\omega 12 = -\bar{e}\cdot\Omega_{12}\cdot\dot{e} = 0.045; \quad \omega 21 = -\bar{e}\cdot\Omega_{21}\cdot\dot{e} = 0.045.$$

Частоты подмножеств $S1$ и $S2$:

$$\omega 1 = -\bar{e}\cdot\Omega_{11}\cdot\dot{e} = 0.045; \quad \omega 2 = -\bar{e}\cdot\Omega_{22}\cdot\dot{e} = 0.045.$$

Матрица частот укрупнённого множества:

$$\Omega_{\text{укр}} = \left\|\omega IJ\right\| = \begin{bmatrix} \omega 11 & \omega 12 \\ \omega 21 & \omega 22 \end{bmatrix} = \begin{bmatrix} -0.045 & 0.045 \\ 0.045 & -0.045 \end{bmatrix}.$$

Видно, что свойство баланса частот справедливо как для исходного множества, так и для укрупнённого множества состояний. Так, частота перехода $S2 \to S1$, вычисленная непосредственно по предельной вероятности состояния и интенсивности перехода:

$$\omega 21 = \pi_5\cdot\lambda_{51} = 0.09\cdot 0.5 = 0.045.$$

Полученные результаты полностью согласуются с результатами модели [6], основанной на использовании теории полумарковских процессов. Так, среднее время цикла при принятых исходных данных составляет 1/0.045 = 22.2 часа, то есть восстановление линии связи происходит со средней периодичностью 22.2 часа. Стационарный коэффициент неготовности, который учитывает время в неработоспособном состоянии линии связи при условии, что она используется по назначению, то есть функционирует:

$$K_{\text{нст}} = \frac{\pi_2}{\pi_1 + \pi_2} = 2\cdot10^{-6}.$$

Заключение

Предложенный частотный метод укрупнения позволяет привести эргодический процесс, состояния которого разбиты на несколько подмножеств по некоторому признаку, к процессу с меньшим числом состояний, при этом информация об исходном множестве состояний определенным образом «упаковывается» в процесс с меньшим числом состояний. В изложенном подходе приведены зависимости между исходными характеристиками процесса, предельными вероятностями состояний исходного и укрупнённого процесса и частотами переходов между состояниями укрупнённого процесса. Особенностью метода является сохранение значений предельных вероятностей подмножеств состояний, средних продолжительностей нахождения в подмножествах состояний и частот попаданий в подмножества состояний соответственно для процесса в дискретном и непрерывном времени. Полученные зависимости между характеристиками исходного и укрупнённого случайного процесса можно использовать при исследовании различных процессов, моделирующих в стационарном режиме функционирование систем различного назначения (технических, экономических, организационных). В частности, по известным (заданным) характеристикам могут быть вычислены другие характеристики и параметры случайных процессов. Например, заданными могут быть переходные вероятности для процесса в дискретном времени или интенсивности переходов в непрерывном времени, частоты состояний и частоты переходов между состояниями как исходного, так и укрупнённого процесса, оценки которых получены путём сбора и обработки информации о случайном процессе. Также в качестве исходных характеристик могут быть использованы продолжительности нахождения в состояниях и в подмножествах состояний, предельные вероятности состояний и подмножеств состояний и т.д. Материал настоящей статьи позволяет составить модель с целью исследования случайного процесса или вероятностной системы на этапе проектирования или совершенствования их эксплуатационных свойств, например, с целью улучшения их эффективности и надёжности.

Характеристики системы находятся путем оперирования с матрицами. При разбиении множества состояний на подмножества происходит переход к блочным матрицам, а при замене подмножеств состояний состояниями укрупнённого процесса – к матрицам меньшего размера по сравнению с исходными. Эффект от укрупнения заключается в том, что уменьшаются размеры матричных характеристик, описывающих переходы между состояниями укрупнённой системы. Таким образом, для нахождения характеристик, связанных с подмножеством состояний, не нужно использовать полные матрицы исходных характеристик, можно оперировать с подматрицами, относящимися к этому подмножеству. В некоторых случаях это может существенно упростить решение конкретных задач.

При моделировании систем могут быть разные цели, разные исходные данные и разные выходные характеристики. На основе полученных соотношений между параметрами исходного и укрупнённого процесса могут быть найдены характеристики, описывающие функционирование моделируемой системы на укрупнённом множестве состояний. Например, при исследовании надёжности систем могут быть найдены коэффициент готовности, среднее время безотказной работы и другие показатели надёжности системы в стационарном режиме.

Алгоритм укрупнения на основе оперирования с матрицами может быть реализован с помощью таких инструментов, как Mathcad или MATLAB. Операции могут быть выполнены в числовом или аналитическом виде.

Литература

1. *Захаров В. К., Сарманов О. В.* Укрупнение состояний цепи Маркова и стационарное изменение спектра // Докл. АН СССР. Физика, математика. 1965. В. 160, № 4. С. 762–764.
2. *Захаров В. М., Эминов Б. Ф.* Алгоритмы укрупнения цепей Маркова // Вестник Казанского гос. технического университета им. А. Н. Туполева. 2013. № 2. С. 125–133.
3. *Зеленцов Б. П.* Матричные модели функционирования оборудования систем связи // Вестник СибГУТИ. 2015. № 4. С. 62–73.
4. *Зеленцов Б. П.* Укрупнение состояний сложных систем, моделируемых марковскими процессами // Вестник СибГУТИ. 2017. № 3. С. 43–56.
5. *Зеленцов Б. П.* Частотный метод моделирования вероятностных систем длительного использования // Вестник СибГУТИ. 2016. № 4. С. 25–38.
6. *Зеленцов Б. П., Максимов В. П., Шувалов В. П.* Модель функционирования линии связи в условиях недостоверного контроля технического состояния // Вестник СибГУТИ. 2015. № 3. С. 35–43.
7. *Каштанов В. А., Медведев А. И.* Теория надежности сложных систем. М.: Физматлит, 2010. 608 с.
8. *Кемени Д., Снелл Д.* Конечные цепи Маркова. М.: Наука, 1970. 272 с.
9. *Клемин А. И., Емельянов В. С., Морозов В. Е.* Расчёт надёжности ядерных энергетических установок. М.: Энергоиздат, 1982. 208 с.
10. *Трофимов А. С., Зеленцов Б. П.* Модель функционирования релейной защиты энергосистем // Электроэнергия. Передача и распределение. 2016. № 6. С. 110–114.

11. *Черкасов А. В.* Принцип квазиэквивалентности укрупнения состояний марковских моделей // Молодой учёный. 2016. № 11. С. 529–535.
12. *Gambin A., Pokarowski P.* A New Combinatorial Algorithm for Large Markov Chains // Computer Algebra in Scientific Computing (CASC 2001). Springer, 2001. P. 195–212.
13. *Gambin A., Pokarowski P.* Aggregation Algorithms for Markov Chains with Large State Space // Institute of Informatics, Institute of Applied Mathematics, Warsaw University, Poland. 49 p.
14. *Ganguly A., Petrov T., Koeppl H.* Markov Chain Aggregation and its Applications to Combinatorial Reaction Networks // arXiv: 1303.4532v2. 2013. 29 p.
15. *Kumar A.* Discrete Event Stochastic Processes. Lecture Notes for Engineering Curriculum, 2012. 166 p.
16. *Rusconi S., Akhmatskaya E., Sokolovski D., Ballard N., de la Cal J.C.* Relative Frequencies of Constrained Events in Stochastic Processes: An analytic approach // Physical Revew E 92. 043306, 2015.
17. *Salfner F.* Modeling Event-driven Time Series with Generalized Hidden Semi-Markov Models. Technical Report 208, Department of Computer Science, Humbold University, 2006.
18. *Shalizi R. C.* Methods and Techniques of Complex System: An overview. Publisher arXiv 2006. 96 p.

Статья поступила в редакцию 04.10.2017;
переработанный вариант – 08.12.2017.

Зеленцов Борис Павлович
д.т.н., профессор, профессор кафедры высшей математики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: zelentsovb@mail.ru.

Aggregation of Markov process states based on frequencies

B. P. Zelentsov

A method of Markov process states aggregation in discrete and continuous time is considered. The initial set of states is partitioned into several subsets. The number of states is reduced to a smaller number of states by replacing each subset by one state of aggregated process. The frequencies of transitions between the states of the initial set are replaced by frequencies of transitions between the aggregated states. The method allows to operate with matrices of smaller dimensions. The method can be applied for systems modeling with a large state space.

Keywords: discrete Markov process in discrete and continuous time, mean frequencies of transitions between the aggregated states, mean frequency of a subset of states.