УДК 519.6

О НОРМАЛЬНОЙ ФОРМЕ СТОХАСТИЧЕСКОЙ МАТРИЦЫ

Ю.А. Альпин, В.С. Альпина

Аннотация

Определяется нормальная форма стохастической матрицы, основанная на понятиях совместимости и целости состояний соответствующей цепи Маркова. В сравнении с классической нормальной формой она содержит больше информации для определения асимптотической эквивалентности состояний.

Ключевые слова: стохастическая матрица, нормальная форма, цепь Маркова.

1. Введение и необходимые сведения

Классификация состояний цепи Маркова, введенная Колмогоровым [1] для анализа асимптотического поведения переходных вероятностей, основана на понятиях о возвратных и невозвратных состояниях, классах сообщающихся состояний и их циклических подклассах. В случае конечной цепи эти понятия приводят к нормальным формам разложимой и неразложимой стохастической матрицы. В настоящей работе предлагаются классификация состояний и нормальные формы стохастической матрицы, основанные на понятиях совместимости и целости состояний. Эта классификация, как и классическая теория, использует лишь комбинаторные свойства матрицы, но даёт больше информации о том, когда два состояния конечной цепи Маркова асимптотически эквивалентны как начальные состояния процесса.

Необходимые для дальнейшего определения и факты, приводимые ниже, более или менее общеизвестны, но для них трудно указать один источник. Помимо основополагающей статьи [1] укажем лишь книги [2–4], из которых можно получить информацию о спектрах, нормальных формах и графах неотрицательных матриц, книгу [5] по матричной теории цепей Маркова, современное изложение теории см. в [6].

Неотрицательная матрица $P = (p_{ij})$ называется стохастической, если суммы элементов каждой из строк равны единице. В теории конечных однородных цепей Маркова число p_{ij} понимается как вероятность перехода цепи из состояния i в состояние j за один такт времени (за один шаг). Вероятности перехода за k шагов даются матрицей $P^k = (p_{ij}^{(k)})$. Поскольку мы не фиксируем начальное состояние (или начальное распределение вероятностей состояний), а исследуем совокупность переходных вероятностей, то будем говорить не о марковской цепи, а о марковской системе, определяемой стохастической матрицей.

Графом неотрицательной матрицы $P = (p_{ij})$ порядка n называют ориентированный граф с вершинами $1, \ldots, n$, в котором дуга ведёт из вершины i в вершину j (коротко: $i \to j$), если $p_{ij} > 0$. Путь $i_1 \ldots i_k i_{k+1}$ длины k называется *простым*, если ни одна его вершина не повторяется, кроме, может быть, первой и последней. При $i_1 = i_{k+1}$ простой путь называется *простым контуром.* Основное соотношение, связывающее неотрицательную матрицу $P = (p_{ij})$ с её графом, заключается в следующем.

Предложение 1. *Элемент* $p_{ij}^{(k)}$ *матрицы* P^k *положителен тогда и только тогда, когда в графе матрицы* P *существует путь длины* k *из вершины* i *в вершину* j.

Говорят, что вершина j *достижима* из вершины i, если $i = j$ или существует (i, j)-путь. При $i \neq j$ всегда можно выбрать простой (i, j)-путь, длина которого не больше чем $n - 1$. О двух вершинах, каждая из которых достижима из другой, говорят, что они *сообщаются*. Если любые две вершины сообщаются, то граф называется *сильно связным*.

Обозначим через B матрицу смежности графа матрицы P. Матрицу B можно определить и как *портрет* матрицы A. Так называется (0,1)-матрица, получаемая из A заменой ненулевых элементов на единицы. Будем считать, что B – матрица над двухэлементной булевой алгеброй. Мы не вводим специальных обозначений для операций булевой алгебры, так как их использование будет ясно из контекста. Например, символ I и в булевском случае обозначает единичную матрицу. Портреты неотрицательных матриц замечательны тем, что портрет суммы (произведения) матриц равен сумме (произведению) их портретов. Следовательно, чтобы вычислить портрет матрицы, образованной путём сложения и умножения неотрицательных матриц, с несравненно меньшими вычислительными затратами можно вместо самих матриц оперировать с их портретами. Например, предложение 1 можно дополнить следующим утверждением:

$$p_{ij}^{(k)} > 0 \;\Leftrightarrow\; b_{ij}^{(k)} = 1.$$

Таким образом, марковская система определяется стохастической матрицей, но если достаточно знать лишь возможность или невозможность перехода за k шагов из состояния i в состояние j, то система с нужной степенью точности определяется орграфом (при этом состояния системы отождествляются с вершинами графа) или, когда нужен аналитический инструмент, – булевой матрицей (иногда в этой ситуации говорят о топологической цепи Маркова).

Состояние марковской системы называется *возвратным*, если оно достижимо из любого достижимого из него состояния. Из возвратного состояния достижимы лишь возвратные, из любого невозвратного состояния достижимо некоторое возвратное состояние.

Множество состояний называется *замкнутым*, если дуги из его состояний ведут лишь в состояния этого множества. Замкнутое множество S определяет *компоненту* системы, которую можно рассматривать как самостоятельную систему. Матрица перехода этой системы есть главная подматрица матрицы P, стоящая на пересечении строк и столбцов с номерами из S. Слово «компонента» будет обозначать и замкнутое множество состояний и определяемую им систему. Компонента называется *разложимой*, если она содержит собственную компоненту. Ясно, что любая компонента содержит неразложимую компоненту. Отметим ещё, что неразложимая компонента состоит из сообщающихся возвратных состояний и всякое возвратное состояние принадлежит некоторой неразложимой компоненте. Различные неразложимые компоненты не пересекаются.

Неотрицательную матрицу P называют *разложимой*, если некоторая перестановка рядов (преобразование подобия, соединяющее перестановку строк с такой же перестановкой столбцов) приводит P к блочно-треугольному виду

$$\begin{pmatrix} P_1 & 0 \\ P_{21} & P_2 \end{pmatrix}, \tag{1}$$

где P_1, P_2 – квадратные матрицы. В противном случае говорят, что P – *неразложимая* матрица. Известно, что неразложимость матрицы равносильна сильной связности её графа.

Применительно к марковским системам перестановка рядов отражает перенумерацию состояний: перестановка i-й и j-й строк, а также i-го и j-го столбцов соответствует тому, что состояние j получает в новой нумерации номер i. Граф данной матрицы и граф матрицы, полученной перестановкой рядов, изоморфны. Новая матрица и её граф задают ту же марковскую систему, но с другой нумерацией состояний. Ситуация здесь аналогична той, что имеет место для линейного оператора и его матриц в различных базисах пространства. «Правильная» нумерация состояний отражает внутреннюю структуру системы и приводит к матрицам нормальной (в том или ином смысле) формы.

Если стохастическая матрица P разложима, то перестановкой рядов она приводится к блочно-треугольной *нормальной форме*

$$\begin{pmatrix} P_1 & & & \\ & \ddots & & \\ & & P_h & \\ P_{h+1,1} & \dots & P_{h+1,h} & P_{h+1,h+1} \end{pmatrix}. \tag{2}$$

Здесь блоки $P_1, \dots, P_h$ – неразложимые стохастические матрицы, отвечающие неразложимым компонентам, матрицы $P_{h+1,1}, \dots, P_{h+1,h}$ содержат вероятности перехода из невозвратных состояний в неразложимые компоненты, матрица $P_{h+1,h+1}$ задаёт переходы внутри множества невозвратных состояний.

Пусть теперь P – неразложимая неотрицательная матрица и ρ – её спектральный радиус. Согласно теореме Фробениуса [2, с. 355] максимальные по модулю собственные значения матрицы P образуют систему

$$\rho, \rho\varepsilon, \dots, \rho\varepsilon^{d-1}, \quad \varepsilon = e^{2\pi i/d}. \tag{3}$$

Число d собственных значений максимального модуля называется *индексом импримитивности* матрицы P. Если $d = 1$, то матрица P *примитивна*, это значит, что при некотором k матрица P^k не содержит нулей. Фробениус открыл связь между индексом импримитивности и расположением ненулевых элементов: неразложимая матрица P с индексом импримитивности $d > 1$ подходящей перестановкой рядов преобразуется к виду

$$\begin{pmatrix} 0 & P_{12} & 0 & \dots & 0 \\ 0 & 0 & P_{23} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & P_{d-1,d} \\ P_{d1} & 0 & 0 & 0 & 0 \end{pmatrix} \tag{4}$$

с квадратными диагональными блоками. Матрица (4) называется нормальной *формой Фробениуса* неотрицательной неразложимой матрицы.

К формам (2) и (4) независимо от аргументов Фробениуса приводит классификация Колмогорова состояний цепи Маркова, основанная на понятиях возвратного и невозвратного состояний, классов сообщающихся состояний и дальнейшего разбиения этих классов на циклические подклассы. Эта классификация не обращается к спектральным свойствам матрицы, а определяется расположением ненулевых элементов в матрице переходных вероятностей, то есть, фактически, графом матрицы.

Оба подхода к форме (4) связывает теорема Романовского [5, с. 102], утверждающая (в иных терминах), что индекс импримитивности неразложимой матрицы равен наибольшему общему делителю длин контуров её графа. Это число будем называть *индексом* графа. Ради краткости будем говорить об индексе матрицы или системы, имея ввиду индекс соответствующего графа.

Вследствие теоремы Романовского форма Фробениуса неразложимой матрицы определяется в чисто комбинаторных терминах – как *блочно-циклическая* матрица вида (4), где d – индекс графа матрицы.

Напомним, что цепь Маркова с матрицей $P = (p_{ij})$ переходных вероятностей называется *эргодической*, если для любых состояний i, j существует предельная вероятность, не зависящая от i:

$$\lim_{k\to\infty} p_{ij}^{(k)} = p_j. \tag{5}$$

Матрицу эргодической цепи (иногда и саму цепь) называют *регулярной*. Надо заметить, что определение эргодичности и регулярности у разных авторов различно. В некоторых источниках (например, в [6]) дополнительно требуется, чтобы предельные вероятности p_j были положительными. При таком определении необходимым и достаточным условием регулярности является примитивность. В других работах (например, в [5]) при определении эргодического свойства допускаются нулевые предельные вероятности. Мы приняли это более широкое определение. Одно из ранних условий регулярности [5, с. 66] гласит:

Предложение 2. *Стохастическая матрица P регулярна тогда и только тогда, когда некоторая её степень P^k содержит положительный столбец.*

В работе Добрушина [7] был введён *коэффициент эргодичности* $\alpha(P) = \min_{i_1,i_2} \sum_j \min(p_{i_1 j}, p_{i_2 j})$ и выяснилось, в частности, что предыдущее условие можно заменить на более слабое: *стохастическая матрица $P = (p_{ij})$ регулярна тогда и только тогда, когда при некотором показателе k*

$$\alpha(P^k) = \min_{i_1,i_2} \sum_j \min(p_{i_1 j}^{(k)}, p_{i_2 j}^{(k)}) > 0.$$

Переформулируем этот критерий на языке графа матрицы.

Предложение 3. *Стохастическая матрица P регулярна в точности тогда, когда существует такое число k, что в графе P из любых двух вершин можно перейти в одну и ту же вершину путями длины k.*

2. Совместимость состояний. Целые состояния

Известно ещё одно определение эргодичности, эквивалентное предыдущему: цепь Маркова с n состояниями и матрицей перехода P называется эргодической, если для любых состояний i_1, i_2

$$\lim_{k\to\infty} (p_{i_1 j}^{(k)} - p_{i_2 j}^{(k)}) = 0, \quad j = 1, \dots, n. \tag{6}$$

Рассмотрим марковскую систему общего вида. Назовём состояния i_1, i_2 *асимптотически эквивалентными*, если для этих состояний выполняется условие (6).

Из классических результатов о предельном поведении переходных вероятностей (см.,например, [6, гл. VIII]) с очевидностью следует, что два возвратных состояния асимптотически эквивалентны в точности тогда, когда они принадлежат одному циклическому подклассу. Но если хотя бы одно из состояний невозвратно, то, вообще говоря, в терминах традиционной классификации вопрос об их асимптотической эквивалентности решить невозможно. Легко привести пример двух марковских систем, у которых графы совпадают, но в первой системе некоторые состояния i_1

и i_2, одно из которых возвратно, а другое невозвратно, асимптотически эквивалентны, а во второй – неэквивалентны. Аналогичный пример легко привести и для пары невозвратных состояний.

Опишем классификацию состояний марковской системы, отличающуюся от традиционной тем, что она даёт – по-прежнему в терминах графа матрицы – необходимое и достаточное условие асимптотической эквивалентности как пары возвратных состояний, так и состояний, лишь одно из которых возвратно. Первые шаги в развиваемом здесь направлении были сделаны, по-видимому, в заметке [8], однако понятия целого состояния и целой стохастической матрицы не были сформулированы и задача построения нормальной формы на их основе не ставилась. Понятие совместимости, определяемое ниже, использовалось в [9] для нового вывода формы Фробениуса неразложимой неотрицательной матрицы.

Введём основные определения. Состояния i_1 и i_2 марковской системы с матрицей P переходных вероятностей *совместимы*, если существует такой показатель k, что для некоторого j одновременно $p_{i_1 j}^{(k)} > 0$ и $p_{i_2 j}^{(k)} > 0$. Иначе говоря, состояния i_1 и i_2 совместимы, если из этих состояний *синхронно*, то есть путями одинаковой длины, достижимо некоторое общее состояние. Желая указать длину путей, будем писать, что состояния k-совместимы. Назовём состояние системы *целым*, если любые синхронно достижимые из него состояния совместимы.

Бинарное отношение совместимости на множестве состояний, очевидно, рефлексивно и симметрично, но в общем случае не транзитивно. Однако системы, на состояниях которых совместимость транзитивна (значит, является отношением эквивалентности), играют главную роль и в общем случае.

Разбиение множества состояний системы называется *циклическим*, если существует такая нумерация классов разбиения: $S_1, S_2, \ldots, S_m$, что все дуги с началом в S_1 ведут в S_2, дуги с началом в S_2 ведут в S_3 и так далее, наконец, из S_m все дуги ведут в S_1. Назовём марковскую систему *целой*, если отношение совместимости на её состояниях является эквивалентностью, классы которой образуют циклическое разбиение. Стохастическую матрицу целой системы будем называть *целой*. Подробнее целые матрицы рассмотрены в следующем параграфе.

Важнейший пример целой системы – неразложимая марковская система, соответственно, неразложимая матрица – пример целой матрицы. Разбиение множества состояний на циклические подклассы является, конечно, циклическим. При этом, как нетрудно видеть, состояния совместимы в точности тогда, когда они принадлежат одному циклическому подклассу, то есть классы совместимости совпадают с циклическими подклассами.

В целой системе пути одинаковой длины, имеющие одно начало, ведут в один класс совместимости, поэтому *все состояния целой марковской системы – целые.*

В частности, целыми являются состояния неразложимых компонент марковской системы. А поскольку любое возвратное состояние принадлежит некоторой неразложимой компоненте, то *все возвратные состояния марковской системы – целые.*

В силу последнего утверждения нецелое состояние невозвратно, поэтому из него достижимо возвратное состояние. В действительности нецелые состояния обладают более сильным свойством:

Предложение 4. *Из нецелого состояния синхронно достижимы два несовместимых возвратных состояния. Такие состояния k-достижимы при любом*

$$k \geq \nu = \frac{1}{2}n(n-1).$$

Доказательство. Пусть i – нецелое состояние, следовательно, имеются пути одинаковой длины, ведущие из i в некоторые несовместимые состояния p и q.

Если оба они возвратные, то доказываемое уже имеется. Если нет, то воспользуемся простым фактом: из любого состояния есть путь длины $n-1$ в некоторое возвратное состояние. Продолжим имеющиеся пути с началом в i путями длины $n-1$ из p и q в некоторые возвратные состояния u и v соответственно. Эти состояния синхронно достижимы из i и несовместимы (совместимость противоречила бы несовместимости p и q). Первое утверждение предложения доказано.

Теперь пусть даны пути наименьшей длины k_0, ведущие из нецелого состояния i в несовместимые возвратные состояния: $ii_1 \dots i_{k_0-1}u$ и $ij_1 \dots j_{k_0-1}v$. Тогда в последовательности пар различных вершин $(i_1, j_1), \dots, (u, v)$ не должно быть совпадающих пар, а также пар, отличающихся лишь порядком элементов. В противном случае данные пути было бы легко перестроить в пути из i в те же состояния u и v одинаковой длины, меньшей, чем k_0. Следовательно, число k_0 не больше, чем количество $\nu = (1/2)n(n-1)$ неупорядоченных пар состояний. Пути длины k_0, ведущие из i в несовместимые возвратные состояния, можно продолжить любыми путями одинаковой длины. Эти продолжения лежат в множестве возвратных состояний и заканчиваются несовместимыми состояниями. □

Аналогично предложению 4 доказывается

Предложение 5. *Если два состояния системы совместимы, то они k-совместимы при любом $k \geq \nu$.*

Из предложений 3 и 5 вытекает

Следствие 1. *Регулярными являются в точности такие системы, у которых любые два состояния совместимы.*

Лемма 1. *Целые состояния образуют замкнутое множество.*

Доказательство. Пусть i – целое и $i \to j$. Если состояния p и q k-достижимы из j, то они $(k+1)$-достижимы из i и поэтому совместимы. Следовательно, j также целое. □

Лемма 2. *Для марковской системы следующие условия равносильны:*
1) *отношение совместимости на состояниях системы транзитивно;*
2) *система целая или является дизъюнктным объединением целых систем;*
3) *все состояния системы целые.*

Доказательство. Докажем, что из 1) следует 2). В рассматриваемом случае совместимость является отношением эквивалентности и определяет разбиение множества состояний на классы. Классы совместимости обладают следующими свойствами.

а) Дуги с началом в одном классе ведут в один и тот же класс.

Действительно, пусть состояния i_1 и i_2 принадлежат классу L. В силу их совместимости есть дуги, ведущие из i_1 и i_2 в некоторый класс M. Но поскольку из целого состояния все дуги ведут в один класс совместимости, то *все* дуги из i_1 ведут в M и *все* дуги из i_2 ведут в M.

б) Дуги, ведущие в один класс, имеют и начало в одном классе.

В самом деле, пусть i_1 и i_2 совместимы, $l_1 \to i_1$, $l_2 \to i_2$,. Тогда, очевидно, l_1 и l_2 также совместимы, то есть лежат в одном классе.

В силу свойства а) на множестве классов совместимости целых состояний корректно определяется отображение, при котором $L \mapsto M$, если дуги из класса L ведут в класс M. Из свойства б) следует, что это отображение – перестановка на множестве классов. Как известно, перестановка, заданная на конечном множестве, разлагается в произведение циклов, другими словами, граф перестановки является объединением непересекающихся простых контуров. Это и значит, что выполнено условие 2).

Как установлено выше, состояния целой системы – целые, поэтому 2) влечёт 3).

Проверим, что из 3) следует 1). Действительно, пусть состояния i_1, i_2, i_3 таковы, что i_1 совместимо с i_2, а i_2 совместимо с i_3. Докажем, что совместимы i_1 и i_3. Согласно предложению 5 из вершин i_1 и i_2 ν-достижима некоторая вершина j_1, а из вершин i_2 и i_3 ν-достижима некоторая вершина j_2. Вершины j_1 и j_2 совместимы, поскольку они синхронно достижимы из целой вершины i_2. Но тогда вершины i_1 и i_3, из которых синхронно достижимы совместимые вершины j_1 и j_2, также совместимы. Транзитивность доказана и доказательство леммы закончено. □

Теорема 1. *Любая стохастическая матрица P либо является целой, либо перестановкой рядов может быть приведена к блочно-треугольному виду*

$$\begin{pmatrix} Q_1 & & & \\ & \ddots & & \\ & & Q_l & \\ Q_{l+1,1} & \dots & Q_{l+1,l} & Q_{l+1} \end{pmatrix}, \tag{7}$$

где $Q_1, \dots, Q_l$ – целые стохастические матрицы, они отвечают целым компонентам марковской системы, определяемой матрицей P; матрицы $Q_{l+1,1}, \dots, Q_{l+1,l}$ содержат вероятности перехода из нецелых состояний в целые компоненты; матрица Q_{l+1} задаёт переходы внутри множества нецелых состояний.

Доказательство. По лемме 1 множество целых состояний образует компоненту. В силу леммы 2 эта компонента состоит из непересекающихся целых компонент. Пусть $W_1, \dots, W_l$ – целые компоненты, $\overline{W}$ – множество нецелых состояний. Тогда при нумерации состояний, согласованной с указанными классами, матрица системы получает желаемый блочно-треугольный вид (7). □

Матрицу (7) назовём *нормальной формой* нецелой стохастической матрицы.

Покажем, как построить нормальную форму (7), используя булев портрет B стохастической матрицы P. Определим *матрицу совместимости* $C = (c_{ij})$ бинарного отношения совместимости как булеву матрицу, для которой $c_{ij} = 1$ тогда и только тогда, когда состояния i и j совместимы. Чтобы вывести формулу для C, выразим условие совместимости на языке матрицы B и используем предложение 5. Тогда получим: состояния i и j совместимы тогда и только тогда, когда строки i и j матрицы B^ν содержат единицы в общем столбце. Отсюда следует, что матрица совместимости выражается формулой

$$C = B^\nu (B^\nu)^T. \tag{8}$$

С помощью матрицы совместимости легко установить, все ли состояния системы целые. Согласно лемме 2 это имеет место, если отношение совместимости транзитивно. Из теории бинарных отношений известно [10, с. 40], что рефлексивное отношение транзитивно тогда и только тогда, когда булева матрица отношения идемпотентна. В результате получаем следующий критерий.

Предложение 6. *Все состояния системы целые в точности тогда, когда $C^2 = C$.*

Если проверка на транзитивность выдержана, то классы совместимости определяются строками C: класс, содержащий состояние i, есть множество $\{j \mid c_{ij} = 1\}$.

Когда классы совместимости известны, определение целых компонент и построение (в этом случае – блочно-диагональной) нормальной формы (7) не представляет трудности.

Если отношение совместимости нетранзитивно, то существует способ найти целые состояния. Согласно предложению 4 состояние i – целое, если ν-достижимые из i состояния совместимы. Придадим этому критерию формально-матричный характер. Обозначим через $D(i) = (d_{uv}(i))$ булеву матрицу бинарного отношения на множестве состояний «быть ν-достижимыми из состояния i». Очевидно, что $d_{uv}(i) = b_{iu}^{(\nu)} b_{iv}^{(\nu)}$. Теперь, используя адамарово (поэлементное) умножение матриц, условие целости состояния можно сформулировать так:

Предложение 7. *Состояние i – целое тогда и только тогда, когда $D(i) \circ C = D(i)$.*

После того, как множество целых состояний определено, строки и столбцы матрицы системы, отвечающие нецелым состояниям, можно вычеркнуть и, оперируя с оставшейся матрицей, провести вычисление блочно-диагональной части нормальной формы (7) по описанной выше схеме. Как видно, вопросы, связанные с построением нормальной формы (7) стохастической матрицы P, решаются с помощью булева портрета P и производных от него матриц.

3. Целые стохастические матрицы

Как видно из предыдущего параграфа, при построении нормальной формы (7) используется понятие целого состояния, более общее, чем понятие возвратного состояния, и, соответственно, понятие целой матрицы, обобщающее понятие неразложимой матрицы. Покажем, что для целой стохастической матрицы существует аналог формы Фробениуса.

Рассмотрим целую стохастическую систему с матрицей P переходных вероятностей и числом m классов совместимости. Пусть $S_1, S_2, \ldots, S_m$ – нумерация классов совместимости, описанная в определении циклического разбиения. Если перенумеровать состояния системы: вначале состояния из S_1, затем из S_2 и так далее, а затем соответственно переставить ряды матрицы системы, то она приобретёт блочно-циклическую форму

$$\begin{pmatrix} 0 & Q_{12} & 0 & \ldots & 0 \\ 0 & 0 & Q_{23} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & Q_{m-1,m} \\ Q_{m1} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9}$$

При возведении матрицы (9) в степень m получим блочно-диагональную матрицу $G = \mathrm{diag}(G_1, G_2, \ldots, G_m)$, где

$$G_1 = Q_{12}Q_{23}\cdots Q_{m1}, \quad G_2 = Q_{23}\cdots Q_{m1}Q_{12}, \ldots, \quad G_m = Q_{m1}Q_{12}\cdots Q_{m-1,m}. \tag{10}$$

Стохастические матрицы (10) удобно рассматривать как матрицы марковских систем, которые возникают, если исходную систему наблюдать в моменты mk, $k = 1, 2, \ldots$ Назовём эти системы *темпоральными* компонентами исходной системы. Множества состояний темпоральных компонент – это классы совместимости.

Если $i, j \in S_t$, то вероятность перехода темпоральной компоненты с матрицей перехода G_t из i в j за k шагов равна вероятности перехода исходной системы с матрицей (9) из i в j за mk шагов. Переводя этот очевидный матричный факт на язык графов, получаем:

Лемма 3. *Пусть состояния i, j принадлежат одной темпоральной компоненте. Тогда (i, j)-путь длины k в графе компоненты существует тогда и только тогда, когда в графе исходной системы существует (i, j)-путь длины mk.*

Лемма 4. *Матрицы темпоральных компонент регулярны.*

Доказательство. По предложению 5 при достаточно большом k в произвольно взятом классе S_t любые два состояния mk-совместимы, а из леммы 3 следует, что эти состояния k-совместимы как состояния темпоральной компоненты. Согласно следствию 1 заключаем, что G_t – регулярная матрица. □

Лемма 5. *Индекс графа целой системы равен числу классов совместимости.*

Доказательство. Пусть $m = 1$, то есть матрица P системы регулярна. Докажем, что индекс графа регулярной матрицы также равен 1. По предложению 2 существует показатель k, при котором в матрице P^k появляется положительный столбец, пусть с номером i. Легко проверить, что тогда и в P^{k+1} столбец с номером i положителен. В частности, имеем $p_{ii}^{(k)} > 0$ и $p_{ii}^{(k+1)} > 0$. Эти неравенства означают, что через вершину i проходят контуры длины k и $k+1$. Следовательно, наибольший общий делитель длин контуров, то есть индекс графа матрицы P, равен 1.

Теперь пусть $m > 1$. В силу леммы 3 в графе темпоральной компоненты контур длины k существует в точности тогда, когда в графе системы есть контур длины mk. Значит, если k_0 – индекс компоненты, K_0 – индекс системы, то $K_0 = mk_0$. В лемме 4 доказано, что компоненты регулярны, значит, $k_0 = 1$, следовательно, $K_0 = m$.

Резюмируя сказанное выше, получаем для целой стохастической матрицы аналог формы Фробениуса (4). □

Теорема 2. *Целая стохастическая матрица P с индексом m является при $m = 1$ регулярной матрицей, а при $m > 1$ некоторой перестановкой рядов приводится к виду* (9).

Матрицу (9) назовём *нормальной формой* целой стохастической матрицы с индексом m.

Если стохастическая матрица P неразложима, то её нормальная форма (9), как целой матрицы, совпадает с нормальной формой Фробениуса. В этом случае матрицы (10) примитивны. Рассмотрим, как соотносятся целость и неразложимость в общем случае.

Теорема 3. *Целая система содержит единственную неразложимую компоненту. Каждый класс совместимости содержит единственный циклический подкласс этой компоненты.*

Доказательство. Ясно, что система содержит некоторую неразложимую компоненту. Каждый циклический подкласс этой компоненты, будучи множеством попарно совместимых состояний системы, целиком содержится в некотором классе совместимости. Невозможно, чтобы в классе совместимости лежали два циклических подкласса одной компоненты, поскольку состояния из различных циклических подклассов несовместимы. Равно невозможно, чтобы целая система содержала две неразложимых компоненты, ведь тогда окажется, что в одном классе совместимости лежат состояния из различных неразложимых компонент. Но такие состояния, конечно, несовместимы. □

Утверждение теоремы 3 можно отразить в строении матрицы (9). Если нумеровать состояния каждого класса совместимости, начиная с принадлежащих циклическому подклассу, то верхний левый угол в каждом стохастическом блоке формы (9), отвечающий состояниям циклических подклассов, будет блоком формы Фробениуса неприводимой компоненты целой системы.

Теорема 3 приводит к следующей характеристике целой системы.

Предложение 8. *Выполнение следующих условий необходимо и достаточно для того, чтобы марковская система была целой:*

1) *отношение совместимости состояний транзитивно;*
2) *система содержит единственную неразложимую компоненту.*

Доказательство. Необходимость. Первое условие следует из того, что все состояния целой системы – целые, и леммы 2. Второе доказано теоремой 3.

Достаточность. Согласно лемме 2 условие 1) означает, что система является целой либо дизъюнктным объединением нескольких целых компонент. Поскольку каждая целая компонента содержит неразложимую компоненту, то вторая возможность противоречит условию 2). Остаётся заключить, что система целая. ☐

Преобразуем предложение 8 в конструктивный критерий, выраженный на языке булевых матриц. Сначала докажем, что условие 2) эквивалентно следующему: *система содержит состояния, достижимые из всех состояний.* Пусть неразложимая компонента единственна. Любое её состояние i достижимо из любого состояния компоненты. Есть также путь в i из любого невозвратного состояния. Действительно, из невозвратного состояния есть путь в некоторое возвратное состояние. Это последнее лежит в компоненте, содержащей i, значит, есть путь из него в i. Соединяя два пути, получим искомый путь. Наоборот, если в системе есть общедостижимые состояния, то совокупность таких состояний, легко видеть, образует единственную неразложимую компоненту. Вопрос о существовании общедостижимых состояний решается с помощью *матрицы достижимости* $R = (r_{ij})$ ($r_{ij} = 1$ тогда и только тогда, когда из i достижимо j): для этого необходимо и достаточно, чтобы R имела положительный столбец. Привлекая еще предложение 6, получаем желаемый критерий:

Следствие 2. *Стохастическая матрица P является целой тогда и только тогда, когда*

1) *булева матрица совместимости C идемпотентна;*
2) *булева матрица достижимости R содержит положительный столбец.*

Матрица C вычисляется по формуле (8), а R – по известной формуле $R = (I + B)^{n-1}$, поэтому проверка целости P является конечной процедурой.

4. Об асимптотической эквивалентности состояний

Сначала решим вопрос об асимптотической эквивалентности целых состояний.

Теорема 4. *Необходимым и достаточным условием асимптотической эквивалентности целых состояний марковской системы является их совместимость.*

Доказательство. Необходимость условия очевидна, докажем достаточность. Совместимые целые состояния принадлежат одной целой компоненте, поэтому достаточно провести рассуждение для целой системы. Будем считать, что стохастическая матрица P системы находится в нормальной форме (9). По лемме 4 матрицы темпоральных компонент регулярны, следовательно, последовательность $P^{mk} = G^k = \operatorname{diag}(G_1^k, G_2^k, \ldots, G_m^k)$ при $k \to \infty$ сходится к матрице

$$G^\infty = \operatorname{diag}(G_1^\infty, G_2^\infty, \ldots, G_m^\infty), \tag{11}$$

в которой каждый диагональный блок есть стохастическая матрица с одинаковыми строками. Разобьём последовательность P^k на m подпоследовательностей

$$P^{mk+r}, \quad k = 1, 2, \ldots,$$

где $r = 0, 1, \ldots, m-1$. Каждая подпоследовательность имеет свой предел, а именно

$$\lim_{k\to\infty} P^{mk+r} = G^{\infty} P^r. \tag{12}$$

Если состояния i_1 и i_2 совместимы и принадлежат темпоральной компоненте S_t, то i_1-я и i_2-я строки матрицы G^{∞} равны, поскольку равны соответствующие им строки в G_t^{∞}. Но тогда строки с этими номерами равны и в остальных предельных матрицах $G^{\infty}P^r$, $r = 1, \ldots, m-1$. Отсюда для состояний i_1 и i_2 следуют предельные равенства (6), что и требовалось доказать. □

Пользуясь теоремой 4, можно установить асимптотическую эквивалентность не только пары возвратных состояний (это можно сделать в рамках традиционной классификации), но и пары, в которой одно или оба состояния невозвратны, лишь бы эти состояния были целыми.

Теорема 5. *Целое и нецелое состояния марковской системы не могут быть асимптотически эквивалентными.*

Доказательство. Пусть показатель ω делится на индекс любой целой компоненты системы. Тогда из любого целого состояния переход за ω шагов возможен лишь в совместимое с ним состояние. Соответственно, матрицы P^{ω} и $P^{\omega k}$ имеют вид

$$P^{\omega} = \begin{pmatrix} H_1 & & & \\ & \ddots & & \\ & & H_z & \\ R_1 & \ldots & R_z & R^{\omega} \end{pmatrix}, \quad P^{\omega k} = \begin{pmatrix} H_1^k & & & \\ & \ddots & & \\ & & H_z^k & \\ R_1^{(k)} & \ldots & R_z^{(k)} & R^{\omega k} \end{pmatrix}. \tag{13}$$

Матрица H_t^k состоит из вероятностей перехода системы за ωk шагов внутри класса S_t, а матрица $R_t^{(k)}$ – из вероятностей перехода за ωk шагов из нецелых состояний в целые состояния класса S_t. Рассмотрим матрицу $P^{\omega(k+1)}$. Вычисление показывает, что $R_t^{(k+1)} = R_t^{(k)} H_t + R^{\omega k} R_t$. Из этого равенства видно, что строчные суммы в $R_t^{(k+1)}$ не меньше соответствующих строчных сумм в $R_t^{(k)}$, то есть вероятность перехода из нецелого состояния i за ωk шагов в класс S_t при росте k не уменьшается. Следовательно, эта вероятность сходится. Из предложения 4 вытекает, что предельные вероятности положительны для двух различных классов совместимости. Поэтому для любого класса предельная вероятность меньше единицы. Между тем, если допустить, что нецелое состояние i асимптотически эквивалентно целому состоянию из некоторого класса S_t, то сумма элементов строки, отвечающей состоянию i в матрице $R_t^{(k)}$, должна при $k \to \infty$ сходиться к единице. Поскольку это невозможно, теорема доказана. □

В заключение сделаем несколько замечаний.

1. Теоремы 4 и 5 не дают ответа на вопрос об асимптотической эквивалентности двух нецелых состояний. Простой пример показывает, что этот вопрос и не может быть решён в терминах комбинаторной структуры матрицы перехода. Пусть

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ a_1 & a_2 & a_3 & 0 \\ b_1 & b_2 & 0 & b_3 \end{pmatrix} \quad (a_3 < 1, b_3 < 1).$$

Легко проверить, что состояния 3 и 4 асимптотически эквивалентны при условии $\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = 0$. Это условие, конечно, не является свойством графа матрицы P.

2. Отличие описанной выше классификации от традиционной особенно заметно при рассмотрении *детерминированных* марковских систем, вероятности перехода в которых равны 1 или 0. В этом случае циклические подклассы состояний одноэлементны, поэтому на языке традиционной классификации асимптотическая эквивалентность никаких двух состояний не определяется. В то же время очевидно, что все состояния детерминированной системы – целые, причём классы совместимости одноэлементны лишь тогда, когда отображение множества состояний в себя, определяемое матрицей перехода, биективно.

3. Согласно [3, с. 60] неотрицательная блочно-циклическая матрица типа (4) без нулевых рядов тогда и только тогда является неприводимой матрицей в форме Фробениуса, когда матрица $P_{12} \cdots P_{d1}$ примитивна. Для стохастической матрицы верно аналогичное утверждение: матрица (9) является нормальной формой целой матрицы в точности тогда, когда матрица $Q_{12} \cdots Q_{m1}$ регулярна.

Summary

Yu.A. Alpin, V.S. Alpina. On the Normal Form of a Stochastic Matrix.

The normal form of a Markov chain's stochastic matrix is defined based on the concepts of state compatibility and safety. Compared to the classical normal form it contains more information for defining asymptotic equivalence of states.

Key words: stochastic matrix, normal form, Markov chain.

Литература

1. *Колмогоров А.Н.* Цепи Маркова со счётным числом возможных состояний // Бюл. Моск. ун-та. Математика и механика. – 1937. – Т. 1, Вып. 3. – С. 1–16.
2. *Гантмахер Ф.Р.* Теория матриц. – М.: Наука, 1967. – 576 с.
3. *Minc H.* Nonnegative matrices. – N. Y.: Wiley, 1988. – 218 p.
4. *Сачков В.Н., Тараканов В.Н.* Комбинаторика неотрицательных матриц. – М.: ТВП, 2000. – 448 с.
5. *Романовский В.И.* Дискретные цепи Маркова. – М.-Л.: Гостехиздат, 1949. – 436 с.
6. *Ширяев А.Н.* Вероятность: в 2 кн. – М.: МЦНМО, 2004. – Кн. 1. – 520 с.; Кн. 2. – 408 с.
7. *Добрушин Р.Л.* Центральная предельная теорема для неоднородных цепей Маркова // Теория вероятностей и её применения. – 1956. – Т. 1, Вып. 1. – С. 72–88; Вып. 4. – С. 365–425.
8. *Альпина В.С.* Классификация состояний вероятностного автомата, ориентированная на изучение его эргодических свойств // Вероятностные методы и кибернетика. – Казань: Изд-во Казан. ун-та, 1980. – Вып. 17. – С. 15–22.
9. *Альпин Ю.А., Альпина В.С.* Теорема Перрона – Фробениуса: доказательство с помощью цепей Маркова // Зап. науч. семин. ПОМИ. – 2008. – Т. 359. – С. 5–16.
10. *Шрейдер Ю.А.* Равенство, сходство, порядок. – М.: Наука, 1971. – 256 с.

Поступила в редакцию
24.01.12

Альпин Юрий Абдуллович – кандидат физико-математических наук, доцент кафедры алгебры и математической логики Казанского (Приволжского) федерального университета.

E-mail: *Yuri.Alpin@ksu.ru*

Альпина Валентина Сергеевна – старший преподаватель кафедры высшей математики Казанского государственного технологического университета.