E. van Doorn, А. И. Зейфман, Т. Л. Панфилова, Оценки и асимптотика скорости сходимости для процессов рождения и гибели, *Теория вероятн. и ее примен.*, 2009, том 54, выпуск 1, 18–38

DOI: 10.4213/tvp2497

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
Том 54 И ЕЕ ПРИМЕНЕНИЯ Выпуск 1
2009

ВАН ДООРН Э. А.[*], *ЗЕЙФМАН А. И.*[**], *ПАНФИЛОВА Т. Л.*[***]

ОЦЕНКИ И АСИМПТОТИКА СКОРОСТИ СХОДИМОСТИ ДЛЯ ПРОЦЕССОВ РОЖДЕНИЯ И ГИБЕЛИ[1)]

Первая часть статьи носит обзорный характер, в ней с той степенью подробности, которая позволяет применять описываемый подход, приводится основная база метода, разработанного одним из авторов в 1990-е годы для получения оценок и явных представлений для скоростей сходимости процессов рождения и гибели. Во второй части работы представлены новые результаты, полученные применением описанного метода к специальным классам процессов рождения и гибели, связанным с моделями среднего поля и системами обслуживания типа $M/M/N/N{+}R$, и относящиеся к асимптотике скорости сходимости в случае, когда число состояний процесса стремится к бесконечности.

Ключевые слова и фразы: процессы рождения и гибели, скорость сходимости, модель среднего поля.

1. Введение. Задачи, связанные с исследованием скорости сходимости к стационарному распределению для эргодических процессов рождения и гибели (далее ПРГ) в зависимости от интенсивностей процесса (интенсивностей рождения и гибели), изучались в литературе с использованием различных классов методов. Эти методы исследования имеют разную степень общности и, соответственно, в различной степени могут быть применены к решению указанных задач. Так, задача может быть сформулирована в терминах операторной теории, а соответствующие общие результаты можно попытаться применить к процессам рождения и гибели. Такой подход рассмотрен, например, в [4]. Можно использовать технику спектральных представлений переходных вероятностей

[*] Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; e-mail: e.a.vandoorn@utwente.nl

[**] Вологодский государственный педагогический университет, Институт проблем информатики РАН, ВНКЦ ЦЭМИ РАН, ул. С. Орлова, 6, 160035 Вологда, Россия; e-mail: zai@uni-vologda.ac.ru

[***] Вологодский государственный педагогический университет, ул. С. Орлова, 6, 160035 Вологда, Россия.

[1)] Второй и третий авторы поддержаны РФФИ (грант № 06-01-00111) и государственным научным грантом Вологодской области.

процессов рождения и гибели, развиваемую, начиная с работ С. Карлина и Дж. Мак-Грегора [16], [17], и сводящую рассматриваемые задачи к теории ортогональных полиномов. Этот подход применен, например, в [8]. В настоящей работе в качестве отправной точки выбирается система прямых уравнений Колмогорова для вероятностей состояний эргодической марковской цепи с непрерывным временем.

В статье преследуются две цели. Во-первых, это обзор и обсуждение метода исследования скорости сходимости для процессов рождения и гибели, инициированного заметкой [11], разработанного одним из авторов в начале 1990-х гг. (см. [29]–[31]) и развитого затем в совместных с Б. Л. Грановским работах (см. [12]–[14]). Во-вторых, представлен ряд новых результатов, полученных этим методом и связанных с асимптотикой скорости сходимости для некоторых специальных классов процессов рождения и гибели в случае, когда число состояний стремится к бесконечности. Метод основан на двух главных составляющих. Прежде всего, это понятие *логарифмической нормы* квадратной матрицы, введенное одновременно и независимо друг от друга С. М. Лозинским [20] и Г. Дальквистом [6] в качестве средства оценки погрешности численного интегрирования системы обыкновенных дифференциальных уравнений (см. также обзорные статьи [24] и [25]). Второй составляющей является специальное преобразование прямой системы Колмогорова. Будет рассмотрен случай конечной марковской цепи с непрерывным временем для обсуждения методологии без подробного описания технических деталей.

Статья построена следующим образом. Пункт 2 начинается с рассмотрения системы дифференциальных уравнений

$$\mathbf{x}'(t) = B\mathbf{x}(t), \qquad t \geqslant 0, \tag{1}$$

где квадратная матрица $B \equiv (b_{ij})$ *существенно положительна* (т.е. все $b_{ij} \geqslant 0$ при $i \neq j$) и неразложима. Показано применение техники, основанной на использовании логарифмической нормы B, для представления и оценивания собственного значения B с максимальной действительной частью и, соответственно, для представления и оценивания (подходящим образом определенной) нормы $\mathbf{x}(t)$ (теоремы 1–3). В п. 3 результаты п. 2 применяются для системы дифференциальных уравнений, получаемой специальным преобразованием прямой системы Колмогорова для вероятностей состояний конечной марковской цепи. В предположении существенной положительности матрицы системы этот подход позволяет получить представление и оценки скорости сходимости к стационарному распределению, а также соответствующие представление и оценки для (подходящим образом определенного) расстояния между распределением вероятностей состояний в момент $t \geqslant 0$ и стационарным распределением (теоремы 4 и 5). Далее, в п. 4 подходы и утверждения п. 3 применяются к процессам рождения и гибели, результаты для которых удается получить без всяких дополнительных условий (теоремы 7–9).

Выбирая конкретные классы интенсивностей рождения и гибели, мы можем теперь получать явные оценки скорости сходимости. В особых случаях удается даже указать точное значение параметра сходимости. Более общими являются ситуации, в которых удается получить асимптотические оценки (когда число состояний стремится к бесконечности) параметра сходимости. В пп. 5 и 6 изучаются два конкретных класса ПРГ, для которых получение такой асимптотики возможно. А именно, ПРГ, соответствующие *моделям среднего поля*, изучаются в п. 5, а процесс, описывающий число требований в системе обслуживания $M/M/N/N+R$, — в п. 6.

2. Предварительные утверждения. Будем обозначать вектор-столбцы рассматриваемой длины, состоящие из нулей и единиц, соответственно $\mathbf{0}$ и $\mathbf{1}$. Неравенства для векторов и матриц понимаются покомпонентно, верхний индекс ^{T} будет обозначать транспонирование. Матричную норму $\|\cdot\|$ будем предполагать индуцированной соответствующей векторной нормой, т.е. $\|A\| := \max\{\|A\mathbf{a}\|\colon \|\mathbf{a}\| = 1\}$. Тогда $\|A\mathbf{a}\| \leqslant \|A\|\,\|\mathbf{a}\|$ для любых матрицы A и вектора $\mathbf{a}$ подходящих размерностей. Далее, для любой квадратной матрицы A с собственным значением λ имеем $|\lambda| \leqslant \|A\|$.

Если $B \equiv (b_{ij})$ — матрица размера $n \times n$, а $\mathbf{x}(\cdot)$ определяется системой (1), то

$$\mathbf{x}(t) = e^{tB}\mathbf{x}(0), \qquad t \geqslant 0.$$

Хорошо известно (см., например, [25, лемма 1c] или [13, предложение 3]), что

$$\|e^{tB}\| \leqslant e^{t\gamma(B)}, \qquad t \geqslant 0, \tag{2}$$

и, следовательно,

$$\|\mathbf{x}(t)\| \leqslant \|e^{tB}\|\,\|\mathbf{x}(0)\| \leqslant e^{t\gamma(B)}\|\mathbf{x}(0)\|, \qquad t \geqslant 0, \tag{3}$$

где

$$\gamma(B) := \lim_{h\downarrow 0} \frac{\|I + hB\| - 1}{h} \tag{4}$$

— *логарифмическая норма* матрицы B.

Обозначим собственные значения матрицы B через $\lambda_1, \lambda_2, \dots, \lambda_n$ и будем считать, что они расположены в порядке убывания их действительных частей, т.е. $\mathrm{Re}\ \lambda_n \leqslant \mathrm{Re}\ \lambda_{n-1} \leqslant \dots \leqslant \mathrm{Re}\ \lambda_1$. Так как $e^{\lambda t}$ есть собственное значение матрицы e^{tB}, если λ — собственное значение B, то

$$e^{t\,\mathrm{Re}\ \lambda_1} = |e^{\lambda_1 t}| \leqslant \|e^{tB}\|, \qquad t \geqslant 0,$$

и, следовательно, учитывая (2), получаем

$$\mathrm{Re}\ \lambda_1 \leqslant \gamma(B). \tag{5}$$

С другой стороны, используя формулу Гельфанда для спектрального радиуса (см., например, [23, теорема 10.13] или [27, теорема 3.4]), имеем

$$\|e^{kB}\|^{1/k} \to e^{\mathrm{Re}\ \lambda_1} \quad \text{при} \quad k \to \infty,$$

а значит,

$$\mathrm{Re}\ \lambda_1 = \inf\{b \in \mathbf{R}\colon \|e^{tB}\| = O\left(e^{bt}\right) \text{ при } t \to \infty\}.$$

Учитывая, что $\|e^{tB}\| = \max\{\|\mathbf{x}(t)\|\colon \|\mathbf{x}(0)\| = 1\}$, получаем также

$$\mathrm{Re}\ \lambda_1 = \inf\{b \in \mathbf{R}\colon \|\mathbf{x}(t)\| = O\left(e^{bt}\right) \text{ при } t \to \infty \text{ для всех } \mathbf{x}(0)\}. \tag{6}$$

Далее будем предполагать, что $\|\cdot\| = \|\cdot\|_1$, т.е. что $\|\mathbf{a}\| = \sum_i |a_i|$ для любого вектора (столбца) $\mathbf{a} \equiv (a_i)$ и соответственно $\|A\| = \max_j \sum_i |a_{ij}|$ для любой матрицы $A \equiv (a_{ij})$. Отсюда, в частности, получаем

$$\gamma(B) = \max_j \left\{ b_{jj} + \sum_{i \neq j} |b_{ij}| \right\}.$$

Если матрица B существенно положительна, то при достаточно большом $r \in \mathbf{R}$ матрица $B + rI$ неотрицательна, а тогда можно применить известные результаты Перрона–Фробениуса (см., например, [22] или [27]). Из них, в частности, вытекает, что $\lambda_1 + r$, а следовательно, и λ_1 действительны. Полагая

$$c_{\min}(B) := \min_j \sum_i b_{ij} \quad \text{и} \quad c_{\max}(B) := \max_j \sum_i b_{ij}, \tag{7}$$

имеем $\gamma(B) = c_{\max}(B)$, а с учетом (5) и (3) получим $\lambda_1 \leqslant c_{\max}(B)$ и

$$\|\mathbf{x}(t)\| \leqslant e^{tc_{\max}(B)} \|\mathbf{x}(0)\|, \qquad t \geqslant 0,$$

соответственно. Более того, в наших условиях оператор e^{tB} положителен, т.е. $e^{tB}\mathbf{x} \geqslant \mathbf{0}$, если $\mathbf{x} \geqslant \mathbf{0}$ (см., например, [13, предложение 1] или [27, теорема 8.2]). В качестве следствия получаем отсюда

$$\mathbf{x}(0) \geqslant \mathbf{0} \quad \Longrightarrow \quad \|\mathbf{x}(t)\| \geqslant e^{tc_{\min}(B)} \|\mathbf{x}(0)\|, \qquad t \geqslant 0 \tag{8}$$

(см. [13, предложение 2], причем условие $\sum_i b_{ij} \leqslant 0$ не является необходимым). Далее, используя (6), имеем оценку $\lambda_1 \geqslant c_{\min}(B)$. Понятно, что в (8) можно вместо $\mathbf{x}(0) \geqslant \mathbf{0}$ взять $\mathbf{x}(0) \leqslant \mathbf{0}$. В итоге получается следующее утверждение.

Теорема 1. *Пусть B существенно положительна, а $\mathbf{x}(0)$ — произвольный вектор начальных условий. Тогда*

$$c_{\min}(B) \leqslant \lambda_1 \leqslant c_{\max}(B) \tag{9}$$

и

$$e^{tc_{\min}(B)}\|\mathbf{x}(0)\| \leqslant \|\mathbf{x}(t)\| \leqslant e^{tc_{\max}(B)}\|\mathbf{x}(0)\|, \qquad t \geqslant 0, \tag{10}$$

причем левое неравенство в (10) *справедливо при выполнении одного из дополнительных условий* $\mathbf{x}(0) \geqslant \mathbf{0}$ *или* $\mathbf{x}(0) \leqslant \mathbf{0}$.

Система (1) не изменится, если мы заменим $\mathbf{x}(0)$ на $-\mathbf{x}(0)$, а B на $-B$, так что можно попытаться применить те же рассуждения и для $-B$. Однако, поскольку $-B$ *не является* существенно положительной, когда B существенно положительна, можно использовать только рассуждения, не опирающиеся на существенную положительность. Положим

$$s_{\min}(B) := \min_j \Big\{ b_{jj} - \sum_{i\neq j} b_{ij} \Big\} \quad \text{и} \quad s_{\max}(B) := \max_j \Big\{ b_{jj} - \sum_{i\neq j} b_{ij} \Big\}. \tag{11}$$

Замечая, что $-\gamma(-B) = \min_j\{b_{jj} - \sum_{i\neq j} |b_{ij}|\} = s_{\min}(B)$, если B существенно положительна, получаем следующее утверждение.

Теорема 2. *Пусть B существенно положительна, а $\mathbf{x}(0)$ произвольно. Тогда*

$$s_{\min}(B) \leqslant \mathrm{Re}\ \lambda_n \tag{12}$$

и

$$e^{ts_{\min}(B)}\|\mathbf{x}(0)\| \leqslant \|\mathbf{x}(t)\|, \qquad t \geqslant 0. \tag{13}$$

Очевидно, что $s_{\min}(B) \leqslant c_{\min}(B)$, если B существенно положительна, так что оценка (13) становится содержательной только в случае, когда вектор $\mathbf{x}(0)$ имеет координаты разных знаков.

Используя предположение о существенной положительности B, можно несколько улучшить результаты теоремы 1. Отметим вначале, что для произвольной невырожденной матрицы T имеем

$$T\mathbf{x}'(t) = (TBT^{-1})T\mathbf{x}(t), \qquad t \geqslant 0,$$

и, применяя теорему 1 не к B, а к TBT^{-1}, получим

$$e^{tc_{\min}(TBT^{-1})}\|\mathbf{x}(0)\|_T \leqslant \|\mathbf{x}(t)\|_T \leqslant e^{tc_{\max}(TBT^{-1})}\|\mathbf{x}(0)\|_T, \qquad t \geqslant 0, \tag{14}$$

где для $(n\times n)$-матрицы T векторная норма $\|\cdot\|_T$ определяется по формуле

$$\|\mathbf{a}\|_T := \|T\mathbf{a}\|,$$

а для выполнения левого неравенства в (14) достаточно, чтобы выполнялось неравенство $T\mathbf{x}(0) \geqslant \mathbf{0}$ или $T\mathbf{x}(0) \leqslant \mathbf{0}$.

Далее, выберем снова $r \in \mathbf{R}$ настолько большим, чтобы матрица $B + rI$ стала неотрицательной. Тогда в соответствии с теоремой Коллатца–Виланда (см., например, [22, с. 666–669] или [27, теорема 2.9]) имеем для любой диагональной матрицы D с положительными диагональными элементами следующие возможности для спектрального радиуса $\rho(B + rI)$ матрицы $B + rI$:

$$c_{\min}(D(B + rI)D^{-1}) < \rho(B + rI) < c_{\max}(D(B + rI)D^{-1})$$

или

$$c_{\min}(D(B + rI)D^{-1}) = \rho(B + rI) = c_{\max}(D(B + rI)D^{-1}).$$

Более того, равенство имеет место тогда и только тогда, когда $D\mathbf{1}$ есть положительный собственный вектор матрицы $B+rI$, соответствующий собственному значению $\rho(B+rI)$ (единственный с точностью до скалярного множителя). Замечая, что $\rho(B + rI) = \lambda_1 + r$, и заменяя T на D в (14), получаем следующее обобщение теоремы 1.

Теорема 3. *Пусть B существенно положительна, а $\mathbf{x}(0)$ — произвольный ненулевой вектор. Тогда для любой диагональной матрицы D с положительными элементами справедливы неравенства*

$$c_{\min}(DBD^{-1}) \leqslant \lambda_1 \leqslant c_{\max}(DBD^{-1}) \tag{15}$$

и

$$e^{tc_{\min}(DBD^{-1})}\|\mathbf{x}(0)\|_D \leqslant \|\mathbf{x}(t)\|_D \leqslant e^{tc_{\max}(DBD^{-1})}\|\mathbf{x}(0)\|_D, \qquad t \geqslant 0, \tag{16}$$

где левое неравенство в (16) заведомо выполняется, если $\mathbf{x}(0) \geqslant \mathbf{0}$ или $\mathbf{x}(0) \leqslant \mathbf{0}$. Граничные значения в (15) (и, следовательно, в (16), если $\mathbf{x}(0) \geqslant \mathbf{0}$ или $\mathbf{x}(0) \leqslant \mathbf{0}$) достигаются тогда и только тогда, когда $D\mathbf{1}$ является положительным собственным вектором матрицы B, соответствующим собственному значению λ_1.

Отсюда немедленно вытекает такое следствие (см. также [27, задача 8.2.6]).

Следствие 1. *Если B существенно положительна, а D — диагональная матрица, то для любого ненулевого начального условия $\mathbf{x}(0) \geqslant \mathbf{0}$ равенство*

$$\|\mathbf{x}(t)\|_D = e^{\lambda_1 t}\|\mathbf{x}(0)\|_D, \qquad t \geqslant 0, \tag{17}$$

выполняется тогда и только тогда, когда $D\mathbf{1}$ — положительный собственный вектор матрицы B, соответствующий собственному значению λ_1.

Для полноты изложения и удобства дальнейших ссылок отметим также результат, получающийся при замене B на DBD^{-1} и $\mathbf{x}(t)$ на $D\mathbf{x}(t)$ в (1) и (13), а именно, неравенство

$$e^{ts_{\min}(DBD^{-1})}\|\mathbf{x}(0)\|_D \leqslant \|\mathbf{x}(t)\|_D, \qquad t \geqslant 0, \tag{18}$$

справедливое для любого вектора $\mathbf{x}(0)$ и любой диагональной матрицы D с положительными диагональными элементами. С учетом теоремы 3, очевидно, что это неравенство становится содержательным только в случае, когда вектор $\mathbf{x}(0)$ имеет координаты разных знаков, поскольку $s_{\min}(DBD^{-1}) \leqslant c_{\min}(DBD^{-1})$.

З а м е ч а н и е. Некоторое небольшое обобщение теоремы 3 и неравенства (18) можно получить, рассматривая вместо (1) систему

$$\mathbf{y}'(t) = q(t)B\mathbf{y}(t), \qquad t \geqslant 0, \tag{19}$$

где (скалярная) функция $q(\cdot)$ предполагается неотрицательной и локально интегрируемой на $[0,\infty)$. В этом случае получаем

$$\mathbf{y}(t) = e^{\tilde{q}(t)B}\mathbf{y}(0), \qquad t \geqslant 0,$$

где $\widetilde{q}(t) := \int_0^t q(u)\,du$. Полагая $\mathbf{x}(t) := \mathbf{y}(\widetilde{q}^{-1}(t))$, можно свести систему (19) к системе (1), применить (16)–(18) и сформулировать результаты в терминах $\mathbf{y}(\cdot)$. Это можно проделать, заменяя $\mathbf{x}(\cdot)$ на $\mathbf{y}(\cdot)$ в (16)–(18) и записывая $\widetilde{q}(t)$ вместо t в показателе.

Если же B *не является* существенно положительной, можно попытаться найти преобразование подобия $C = TBT^{-1}$ такое, чтобы C стала существенно положительной, а затем применить теорему 3 к C вместо B для получения оценок λ_1.

3. Сходимость марковских цепей. Рассмотрим марковскую цепь с непрерывным временем $\mathscr{X} \equiv \{X(t),\ t \geqslant 0\}$, принимающую значения из множества $\{0,1,\dots,n\}$ с q-матрицей $Q \equiv (q_{ij})$. Вероятности состояний цепи $p_j(t) := \Pr\{X(t) = j\}$, $j = 0,1,\dots,n$, определяются начальным распределением и системой прямых уравнений Колмогорова

$$\mathbf{p}'(t) = Q^T\mathbf{p}(t), \qquad t \geqslant 0, \tag{20}$$

где $\mathbf{p}(t) \equiv (p_0(t), p_1(t), \dots, p_n(t))^T$. Будем предполагать, что цепь $\mathscr{X}$ консервативна и неразложима, а значит, имеет единственное стационарное распределение, определяемое вектором $\boldsymbol{\pi} \equiv (\pi_0, \pi_1, \dots, \pi_n)^T$. Распределение $\boldsymbol{\pi}$ удовлетворяет условию $Q^T\boldsymbol{\pi} = \mathbf{0}$, а $\mathbf{p}(t)$ сходится к $\boldsymbol{\pi}$ при $t \to \infty$ при любом начальном распределении $\mathbf{p}(0)$. Нас будет интересовать параметр сходимости (*decay parameter*) β, определяемый формулой

$$\beta := \sup\{a > 0 : \|\mathbf{p}(t) - \boldsymbol{\pi}\| = O(e^{-at}) \text{ при } t \to \infty \text{ для всех } \mathbf{p}(0)\}. \tag{21}$$

Поскольку $\sum_{k=0}^{n} p_k(t) = \sum_{k=0}^{n} \pi_k = 1$, имеем очевидное неравенство

$$0 \leqslant |p_0(t) - \pi_0| \leqslant \sum_{k=1}^{n} |p_k(t) - \pi_k|, \qquad t \geqslant 0.$$

Теперь, полагая $\mathbf{x}(t) := (p_1(t) - \pi_1, p_2(t) - \pi_2, \ldots, p_n(t) - \pi_n)^T$, приходим к оценке

$$\|\mathbf{x}(t)\| \leqslant \|\mathbf{p}(t) - \boldsymbol{\pi}\| \leqslant 2\|\mathbf{x}(t)\|, \qquad t \geqslant 0, \tag{22}$$

а значит,

$$\beta = \sup\{a > 0\colon \|\mathbf{x}(t)\| = O(e^{-at}) \text{ при } t \to \infty \text{ для всех } \mathbf{x}(0)\}. \tag{23}$$

Отметим также, что из системы (20) вытекает

$$\begin{aligned} p_i'(t) &= q_{0i}(p_0(t) - \pi_0) + \sum_{j=1}^{n} q_{ji}(p_j(t) - \pi_j) \\ &= \sum_{j=1}^{n} (q_{ji} - q_{0i})(p_j(t) - \pi_j), \qquad i = 1, 2, \ldots, n, \quad t \geqslant 0, \end{aligned}$$

так что

$$x_i'(t) = \sum_{j=1}^{n} (q_{ji} - q_{0i}) x_j(t), \qquad i = 1, 2, \ldots, n, \quad t \geqslant 0.$$

Таким образом, $\mathbf{x}(\cdot)$ удовлетворяет системе дифференциальных уравнений (1), а значит, $\mathbf{x}(t) = e^{tB}\mathbf{x}(0)$, где матрица $B \equiv (b_{ij})$ определяется следующим образом:

$$b_{ij} := q_{ji} - q_{0i}, \qquad i, j = 1, 2, \ldots, n. \tag{24}$$

Как и в предыдущем пункте, обозначим собственные значения матрицы B через $\lambda_1, \lambda_2, \ldots, \lambda_n$ и будем считать, что $\mathrm{Re}\ \lambda_n \leqslant \mathrm{Re}\ \lambda_{n-1} \leqslant \cdots \leqslant \mathrm{Re}\ \lambda_1$. Поскольку марковская цепь $\mathscr{X}$ консервативна и неразложима, нулевое собственное значение матрицы Q является простым, а все остальные имеют отрицательные действительные части. Следовательно, $\lambda_1, \lambda_2, \ldots, \lambda_n$ — ненулевые собственные значения Q, причем

$$\mathrm{Re}\ \lambda_n \leqslant \mathrm{Re}\ \lambda_{n-1} \leqslant \cdots \leqslant \mathrm{Re}\ \lambda_1 < 0.$$

В результате, используя (6) и (23), получаем

$$\beta = -\,\mathrm{Re}\ \lambda_1, \tag{25}$$

а значит, изучение скорости сходимости марковской цепи $\mathscr{X}$ с q-матрицей $Q \equiv (q_{ij})$ сводится в значительной степени к исследованию собственного значения с наибольшей действительной частью для матрицы $B \equiv (b_{ij})$, определяемой формулой (24).

Хотя исходная матрица Q существенно положительна, матрица B, вообще говоря, *таковой не является.* А тогда, как упоминалось в конце предыдущего пункта, прежде чем применять теорему 3, надо подобрать преобразование $C = TBT^{-1}$ (если оно, конечно, есть) такое, что C уже является существенно положительной. В следующем пункте мы рассмотрим класс таких преобразований для ПРГ, а пока будем просто предполагать существование. В качестве следствия такого предположения мгновенно получаем действительность λ_1. Очевидно, далее, что если все-таки сама матрица B оказывается существенно положительной, то можно взять $T = I$. Можно еще добавить, что в соответствии с [13, теорема 2] подходящее преобразование существует всегда в случае, если цепь $\mathscr{X}$ является *обратимой* (*reversible*) (в этом случае *все* собственные значения действительны).

В предположении существования невырожденного преобразования T такого, что $C = TBT^{-1}$ существенно положительна, можно применить первую часть теоремы 3 к C и, с учетом (25), получить следующую информацию о β, а значит, и о скорости сходимости марковской цепи $\mathscr{X}$.

Теорема 4. *Пусть T — невырожденная матрица такая, что $C = TBT^{-1}$ существенно положительна, и пусть D — диагональная матрица с положительными диагональными элементами. Тогда*

$$c_{\min}(DT(-B)(DT)^{-1}) \leqslant \beta \leqslant c_{\max}(DT(-B)(DT)^{-1}), \tag{26}$$

причем равенство достигается тогда и только тогда, когда $D\mathbf{1}$ является собственным вектором матрицы C, соответствующим собственному значению $\lambda_1 = -\beta$.

Полагая $\mathbf{z}(\cdot) := T\mathbf{x}(\cdot)$, имеем

$$\mathbf{z}'(t) = TBT^{-1}\mathbf{z}(t) = C\mathbf{z}(t), \qquad t \geqslant 0,$$

причем $\mathbf{x}(\cdot)$ удовлетворяет системе дифференциальных уравнений (1). Отметим, далее, что

$$\|\mathbf{z}(t)\|_D = \|DT\mathbf{x}(t)\| \leqslant \|DT\|\,\|\mathbf{x}(t)\|, \qquad t \geqslant 0,$$

и

$$\|\mathbf{x}(t)\| = \|(DT)^{-1}D\mathbf{z}(t)\| \leqslant \|(DT)^{-1}\|\,\|\mathbf{z}(t)\|_D, \qquad t \geqslant 0,$$

причем с учетом (22)

$$\|DT\|^{-1}\,\|\mathbf{z}(t)\|_D \leqslant \|\mathbf{p}(t) - \boldsymbol{\pi}\| \leqslant 2\,\|(DT)^{-1}\|\,\|\mathbf{z}(t)\|_D, \qquad t \geqslant 0. \tag{27}$$

Теперь можно применить к $\mathbf{z}(\cdot)$ вторую часть теоремы 3 и оценку (18). Используя неравенства (27), можно легко получить следующие явные оценки для $\|\mathbf{p}(t) - \boldsymbol{\pi}\|$.

Теорема 5. *Пусть невырожденная матрица T такова, что $C = TBT^{-1}$ существенно положительна, и пусть D — диагональная матрица с положительными диагональными элементами. Тогда для любого начального распределения $\mathbf{p}(0)$ выполнены неравенства*

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \leqslant \kappa\, e^{-c_{\min} t}\|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0, \tag{28}$$

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \geqslant \frac{1}{\kappa} e^{-s_{\max} t}\|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0, \tag{29}$$

где $\kappa := 2\,\|DT\|\,\|(DT)^{-1}\|$, $c_{\min}$ *и* $c_{\max}$ *обозначают соответственно левую и правую части* (26), *а* $s_{\max} := s_{\max}(DT(-B)(DT)^{-1})$. *Кроме того, если все* $p_i(0) \geqslant \pi_i$, $i = 1, 2, \ldots, n$, *или все* $p_i(0) \leqslant \pi_i$ *при* $i = 1, 2, \ldots, n$, *то*

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \geqslant \frac{1}{\kappa} e^{-c_{\max} t}\|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0. \tag{30}$$

Отметим, что условие (30), в частности, выполнено, если начальное распределение вероятностей сконцентрировано в нуле. Разумеется, оценка (29) становится неэффективной, если (30) выполнено, поскольку $s_{\max} < c_{\max}$.

Поскольку B и $DTB(DT)^{-1}$ имеют одинаковый спектр, теорема 2 гарантирует, что $-s_{\max} \leqslant \operatorname{Re} \lambda_n$. В следующем пункте мы получим более точные оценки для λ_n — собственного значения матрицы Q с минимальной действительной частью в случае ПРГ.

Отметим наконец, что справедливо следующее утверждение.

Теорема 6. *Если для некоторой невырожденной матрицы T матрица $C = TBT^{-1}$ является существенно положительной, то существует неотрицательная матрица Δ такая, что для любого начального распределения $\mathbf{p}(0)$, удовлетворяющего условию $p_i(0) \geqslant \pi_i$ при $i = 1, 2, \ldots, n$ (или $p_i(0) \leqslant \pi_i$ при $i = 1, 2, \ldots, n$),*

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\|_\Delta = e^{-\beta t}\|\mathbf{p}(0) - \boldsymbol{\pi}\|_\Delta, \qquad t \geqslant 0.$$

Доказательство. Выберем диагональную матрицу D так, чтобы $D\mathbf{1}$ был положительным собственным вектором C, соотвествующим λ_1. Тогда из следствия 1, примененного к C и $\mathbf{z}(\cdot)$, получаем

$$\|\mathbf{z}(t)\|_D = e^{-\beta t}\|\mathbf{z}(0)\|_D, \qquad t \geqslant 0.$$

Если Δ — матрица с нулевыми первой строкой и первым столбцом и с матрицей DT в качестве остальной части, то получаем

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\|_\Delta = \|\mathbf{x}(t)\|_{DT} = \|\mathbf{z}(t)\|_D, \qquad t \geqslant 0,$$

откуда и следует требуемое утверждение.

В следующем пункте будет показано, как общая техника, описанная здесь для общего анализа скорости сходимости конечной марковской цепи с непрерывным временем, работает и позволяет получить явные представления и оценки в более специальном случае процессов рождения и гибели.

4. Процессы рождения и гибели. В этом пункте предполагается, что рассматриваемая марковская цепь $\mathscr{X}$ является процессом рождения и гибели (ПРГ), т.е. ее q-матрица имеет вид

$$Q = \begin{pmatrix} -a_0 & a_0 & 0 & \cdots & 0 & 0 \\ b_1 & -(a_1+b_1) & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(a_{n-1}+b_{n-1}) & a_{n-1} \\ 0 & 0 & 0 & \cdots & b_n & -b_n \end{pmatrix} \tag{31}$$

с положительными *интенсивностями рождения* a_i, $0 \leqslant i < n$, и *интенсивностями гибели* b_i, $0 < i \leqslant n$. Как известно, ненулевые собственные значения $\lambda_1, \lambda_2, \ldots, \lambda_n$ матрицы Q действительны, различны и отрицательны (см., например, [17]). В предыдущем пункте было отмечено, что все эти собственные значения являются и точками спектра матрицы B из (24), которая теперь выглядит следующим образом:

$$B = \begin{pmatrix} -(a_0+a_1+b_1) & b_2-a_0 & -a_0 & \cdots & -a_0 & -a_0 \\ a_1 & -(a_2+b_2) & b_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(a_{n-1}+b_{n-1}) & b_n \\ 0 & 0 & 0 & \cdots & a_{n-1} & -b_n \end{pmatrix}.$$

Выбирая в качестве T верхнетреугольную матрицу с единицами на диагонали и над ней, получаем

$$T = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{32}$$

а преобразование подобия $C = TBT^{-1}$ приводит нас к существенно положительной матрице

$$C = \begin{pmatrix} -(a_0+b_1) & b_1 & 0 & \cdots & 0 & 0 \\ a_1 & -(a_1+b_2) & b_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(a_{n-2}+b_{n-1}) & b_{n-1} \\ 0 & 0 & 0 & \cdots & a_{n-1} & -(a_{n-1}+b_n) \end{pmatrix}. \quad (33)$$

Теперь можно применить теорему 4 и после небольших вычислений получить следующее утверждение, которое в полной общности было впервые получено в [31].

Теорема 7. *Параметр сходимости β для ПРГ с q-матрицей* (31) *удовлетворяет условию*

$$\max_{\mathbf{d}>\mathbf{0}} \Big\{ \min_{1\leqslant i\leqslant n} \alpha_i \Big\} = \beta = \min_{\mathbf{d}>\mathbf{0}} \Big\{ \max_{1\leqslant i\leqslant n} \alpha_i \Big\}, \quad (34)$$

где $\mathbf{d} \equiv (d_1, d_2, \ldots, d_n)$ *и*

$$\alpha_i := a_{i-1} + b_i - \frac{d_{i+1}}{d_i} a_i - \frac{d_{i-1}}{d_i} b_{i-1}, \qquad i = 1, 2, \ldots, n, \quad (35)$$

причем $a_n = b_0 = d_0 = d_{n+1} = 0$.

Имеется обширная литература, связанная с результатами, близкими к теореме 7, основанная на различных подходах, а в ряде работ исследован и случай эргодического ПРГ со счетным пространством состояний. Укажем здесь работы [2]–[4], [7]–[9], [12], [13], [15], [18], [29]–[32], см. также приведенные в этих работах ссылки.

Информацию о λ_n — наименьшем собственном значении матрицы Q, а значит, и матрицы C, можно получить, заметив, что $-\lambda_n$ является наибольшим собственным значением матрицы $-C$. Поскольку характеристический полином $P_r(\lambda)$ для главной подматрицы порядка r матрицы $-C$ удовлетворяет рекуррентным формулам

$$\begin{aligned} P_r(\lambda) &= (a_{r-1} + b_r - \lambda) P_{r-1}(\lambda) - a_{r-1} b_{r-1} P_{r-2}(\lambda), \quad r = 2, 3, \ldots, n, \\ P_0(\lambda) &= 1, \quad P_1(\lambda) = a_0 + b_1 - \lambda, \end{aligned}$$

понятно, что $P_n(\lambda)$, а следовательно, и спектр матрицы $-C$ не меняется при изменении знаков ее внедиагональных элементов. Другими словами, $-\lambda_1 = \beta < -\lambda_2 < \cdots < -\lambda_n$ являются собственными значениями неотрицательной (и неразложимой, а значит, существенно положительной) матрицы C^*, которая получается из C изменением знаков ее диагональных элементов. Применяя теперь теорему 3 к C^*, получаем следующее утверждение.

Теорема 8. *Пусть $\chi = -\lambda_n$, где λ_n — наименьшее собственное значение q-матрицы* (31). *Тогда для χ справедливы представления*

$$\max_{\mathbf{d}>\mathbf{0}} \Big\{ \min_{1\leqslant i\leqslant n} \zeta_i \Big\} = \chi = \min_{\mathbf{d}>\mathbf{0}} \Big\{ \max_{1\leqslant i\leqslant n} \zeta_i \Big\}, \quad (36)$$

где $\mathbf{d} \equiv (d_1, d_2, \ldots, d_n)$ *и*

$$\zeta_i := a_{i-1} + b_i + \frac{d_{i+1}}{d_i} a_i + \frac{d_{i-1}}{d_i} b_{i-1}, \qquad i = 1, 2, \ldots, n, \tag{37}$$

причем $a_n = b_0 = d_0 = d_{n+1} = 0$.

Приведем теперь соответствующие экспоненциальные оценки для $\|\mathbf{p}(t) - \boldsymbol{\pi}\|$. Результаты в основном были получены в [29] и [31].

Теорема 9. *Пусть* $\mathscr{X}$ *— ПРГ с* q*-матрицей* (31). *Тогда для любого начального распределения* $\mathbf{p}(0)$ *и любого вектора* $\mathbf{d} \equiv (d_1, d_2, \ldots, d_n)$ *такого, что* $d := \min_i\{d_i\} > 0$, *справедливы оценки*

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \leqslant \kappa\, e^{-\min_i\{\alpha_i\}t} \|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0, \tag{38}$$

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \geqslant \frac{1}{\kappa}\, e^{-\max_i\{\zeta_i\}t} \|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0, \tag{39}$$

где $\kappa = 4\sum_{i=1}^n (d_i/d)$, *а* α_i *и* ζ_i *определяются соответственно формулами* (35) *и* (37). *Кроме того, если* $p_i(0) \geqslant \pi_i$ *при* $i = 1, 2, \ldots, n$ (*или* $p_i(0) \leqslant \pi_i$ *при* $i = 1, 2, \ldots, n$), *то*

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \geqslant \frac{1}{\kappa} e^{-\max_i\{\alpha_i\}t} \|\mathbf{p}(0) - \boldsymbol{\pi}\|, \qquad t \geqslant 0. \tag{40}$$

Доказательство. Полагая $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$ и выбирая T как в (32), получаем

$$\|DT\| = \sum_{i=1}^n d_i \quad \text{и} \quad \|(DT)^{-1}\| = \frac{2}{d}.$$

Оценки (38) и (40) вытекают теперь из теоремы 5, а (39) следует из (18).

Отметим еще, что, в соответствии с теоремой 8, наилучший показатель экспоненты в нижней оценке (39) равен $-\chi t$. В случае произвольного начального распределения (и $n > 1$) этот показатель неулучшаем и несложно построить примеры, для которых $\|\mathbf{p}(t) - \boldsymbol{\pi}\| = O(e^{-\chi t})$ при $t \to \infty$. Однако надо иметь в виду, что при больших значениях t нижняя оценка (39) окажется (почти наверно) слишком грубой, поскольку спектральное представление для величин $p_j(t) - \pi_j$, а значит, и для $\|\mathbf{p}(t) - \boldsymbol{\pi}\|$, содержит и слагаемые порядка $e^{-\beta t}$, так что указанная общая оценка окажется точной при очень специальном выборе интенсивностей и начальных условий (если начальные условия, скажем, выбирать случайно, то с нулевой вероятностью).

В завершение этого пункта рассмотрим специальный пример ПРГ с постоянными (не зависящими от состояния) интенсивностями рождения $a_i = a$, $0 \leqslant i < n$, и гибели $b_i = b$, $0 < i \leqslant n$. Чтобы отметить зависимость от размерности n, будем писать β_n и χ_n вместо β и χ соответственно. Выбирая

$$d_i = \Big(-\sqrt{\frac{b}{a}}\Big)^{i-1} \frac{\sin(in\pi/(n+1))}{\sin(n\pi/(n+1))}, \qquad i = 1, \ldots, n$$

(d_i положительны, так как $(i-1)\pi < in\pi/(n+1) < i\pi$), и используя тождество

$$\sin\frac{(i+1)n\pi}{n+1} + \sin\frac{(i-1)n\pi}{n+1} = 2\sin\frac{in\pi}{n+1}\cos\frac{n\pi}{n+1}, \qquad i = 1, \ldots, n,$$

получаем, что величины α_i в (35) и ζ_i в (37) постоянны и вычисляются по формулам

$$\alpha_i = a + b - 2\sqrt{ab}\cos\frac{\pi}{n+1}, \quad \zeta_i = a + b + 2\sqrt{ab}\cos\frac{\pi}{n+1},$$

$i = 1, \ldots, n$. А тогда в соответствии с теоремами 7 и 8 получаем

$$\beta_n = a+b-2\sqrt{ab}\cos\frac{\pi}{n+1}, \quad \chi_n = a+b+2\sqrt{ab}\cos\frac{\pi}{n+1}, \qquad n \geqslant 1. \tag{41}$$

Эти результаты хорошо известны (см., например [26, с. 13]) и [12, пример 2.3]), однако доказательство представляется новым. Из явных представлений вытекает, что

$$\lim_{n\to\infty} \beta_n = (\sqrt{a} - \sqrt{b})^2 \quad \text{и} \quad \lim_{n\to\infty} \chi_n = (\sqrt{a} + \sqrt{b})^2, \tag{42}$$

и получаем простейшие примеры асимптотических результатов, которые будут изучаться в следующих пунктах.

5. Модели среднего поля. В этом пункте мы рассмотрим спиновую систему $\Phi \equiv (\boldsymbol{\phi}(t),\ t \geqslant 0)$ на полном графе с n узлами (vertices, sites). Пространство состояний такого процесса есть множество всех 2^n бинарных векторов длины n, где i-я компонента вектора состояний принимает значение 0, если i-е состояние свободно, и 1 — если оно занято. Процесс задается с помощью $2n$ неотрицательных *интенсивностей перескока* (*flip rates*) λ_i и μ_i, $i = 0, 1, \ldots, n-1$, где λ_i и μ_i — интенсивности рождения и гибели для свободного (или соответственно занятого) узла, имеющего i занятых соседних узлов. Такие спиновые системы известны в статистической физике как модели среднего поля (*mean-field models*).

Обозначим $|\mathbf{a}|$ общее число единиц бинарного вектора $\mathbf{a}$ (так что $|\boldsymbol{\phi}(t)|$ — общее число занятых узлов в момент t) и положим $X(t) := |\boldsymbol{\phi}(t)|$, $t \geqslant 0$; понятно, что $\mathscr{X} := \{X(t),\ t \geqslant 0\}$ — процесс рождения и гибели с пространством состояний $\{0, 1, \ldots, n\}$ и интенсивностями рождения и гибели $a_i = (n-i)\lambda_i$, $0 \leqslant i < n$, и $b_i = i\mu_{i-1}$, $0 < i \leqslant n$, соответственно. Следовательно, скорость сходимости к стационарному режиму для модели среднего поля Φ характеризуется скоростью сходимости соответствующего ПРГ $\mathscr{X}$, а значит, для исследования сходимости моделей среднего поля можно применять методы предыдущего пункта. Этот подход был развит в работах [12], [13] и [15].

Так, например, с использованием теоремы 7 в [12] было показано, что если $\lambda_i + \mu_i = c$ при $0 \leqslant i \leqslant n-1$, то

$$\beta_n \geqslant c - (n-1)\max_i\{|\lambda_i - \lambda_{i-1}|, |\mu_i - \mu_{i-1}|\}, \qquad n \geqslant 1,$$

причем равенство достигается, если и только если $\lambda_i = \lambda_0 + ih$, $\mu_i = \mu_0 - ih$, $0 \leqslant i \leqslant n-1$, при некотором действительном h, в этом случае $\beta_n = c-(n-1)h$, $n \geqslant 1$. Аналогичным образом в [15] было показано, что при $\lambda_i + \mu_i = c$ значение χ_n находится точно, а именно, $\chi_n = nc$, $n \geqslant 1$. Как и в конце предыдущего пункта, зависимость от n будем учитывать, записывая β_n и χ_n вместо β и χ соответственно.

Специальный класс моделей среднего поля с равномерным стационарным распределением возникает, если $\lambda_i = \mu_i$, $0 \leqslant i \leqslant n-1$. Случай $\lambda_i = \mu_i = c(i+1)^s$, $0 \leqslant i \leqslant n-1$, при некотором $c > 0$ и $s \geqslant 0$ был рассмотрен в [13] и [15]. В частности, используя теорему 7, в [13] было показано, что

$$\begin{aligned} &\lim_{n\to\infty} \frac{\beta_n}{n^s} = 2^{1-s}c, \qquad 0 \leqslant s < 1, \\ &\lim_{n\to\infty} \frac{\beta_n}{n} = c, \qquad s \geqslant 1, \end{aligned} \tag{43}$$

а асимптотика для χ_n была исследована в [15]. Формула (43) допускает обобщение, как показывает следующее утверждение.

Теорема 10. *Пусть* $\lambda_0 = \mu_0 = c > 0$ *и, кроме того,* $\lambda_i = \mu_i \geqslant c(i+1)$ *при* $0 < i \leqslant n-1$. *Тогда*

$$\lim_{n\to\infty} \frac{\beta_n}{n} = c.$$

Доказательство. Записывая $d_1 = d/\lambda_0$ и выбирая $d_{i+1} := 1/(i\lambda_i)$ для $i = 1, 2, \ldots, n$, после небольших преобразований приходим к следующим формулам, определяющим α_i из (35):

$$\alpha_i = \begin{cases} \lambda_0\Big(\Big(1 - \dfrac{1}{d}\Big)n + 1 + \dfrac{1}{d}\Big), & i = 1, \\ \dfrac{\lambda_1}{2}(n + 4 - 2d), & i = 2, \\ \dfrac{\lambda_{i-1}}{i}\Big(n - \dfrac{i}{i-2}\Big), & i = 3, 4, \ldots, n. \end{cases}$$

Теперь, выбирая $d := \sqrt{n}$, получаем, что

$$\frac{\beta_n}{n} \geqslant \min_{1\leqslant i\leqslant n} \left\{\frac{\alpha_i}{n}\right\} \geqslant c(1+o(1)) \quad \text{при} \quad n \to \infty.$$

С другой стороны, имеем

$$\min_{1\leqslant i\leqslant n} \alpha_i \leqslant \alpha_1 = c\left(n+1-\frac{n-1}{d_1}\right) \leqslant c(n+o(n)) \quad \text{при} \quad n \to \infty$$

для любого $d_1 > 0$, так что результат вытекает из минимаксного представления теоремы 7.

6. Система обслуживания $M/M/N/N+R$**.** В этом пункте мы рассмотрим систему обслуживания $M/M/N/N+R$, где $N \geqslant 1$ — число приборов (серверов), а $R \geqslant 0$ — число мест ожидания. Будем предполагать, что интенсивность поступления требований $\lambda(N) > 0$ и число мест ожидания $R \equiv R(N)$ являются функциями от N. Если интенсивность обслуживания одним прибором есть μ, то число требований в системе (длина очереди) есть ПРГ с интенсивностями

$$a_i = \lambda(N), \quad 0 \leqslant i < n(N), \quad \text{и} \quad b_i = \min\{i, N\}\mu, \quad 0 < i \leqslant n(N),$$

где $n(N) := N + R(N)$. Исследуем здесь асимптотическое поведение при $N \to \infty$ параметра сходимости $\beta \equiv \beta_{n(N)}$, а также нижней границы скорости сходимости $\chi \equiv \chi_{n(N)}$. Первый результат относится к ситуации, когда $\lambda(N)$ есть константа.

Теорема 11. *Пусть* $\lambda(N) = \lambda > 0$, *тогда, независимо от поведения* $R(N)$,

$$\lim_{N\to\infty} \beta_{n(N)} = \mu.$$

Д о к а з а т е л ь с т в о. Выбирая

$$d_i = \begin{cases} 1, & 1 \leqslant i \leqslant N, \\ \left(\dfrac{N}{N-1}\right)^{i-N}, & N < i < N+R(N), \\ \dfrac{(N-1)\mu}{(N-1)\mu+\lambda}\left(\dfrac{N}{N-1}\right)^{R(N)}, & i = N+R(N) \ (\text{если } R(N) > 0), \end{cases}$$

получаем после очевидных преобразований следующие равенства для величин α_i из (35):

$$\alpha_i = \begin{cases} \mu, & 1 \leqslant i < N \quad \text{и} \quad i = N+R(N), \\ \mu - \dfrac{\lambda}{N-1}, & N \leqslant i < N+R(N)-1, \\ \mu - \dfrac{\lambda(\lambda-\mu)}{(N-1)\mu+\lambda}, & i = N+R(N)-1 \ (\text{если } R(N) > 1). \end{cases}$$

Следовательно, все α_i равны $\mu + o\,(1)$ при $N \to \infty$, а тогда по теореме 7 получаем требуемое утверждение.

Несложно убедиться, что в условиях теоремы 11 выполняется также и равенство

$$\lim_{N\to\infty} \frac{\chi_{n(N)}}{N} = \mu. \tag{44}$$

Действительно, выбирая $d_i = N^{i/2}$ при $1 \leqslant i \leqslant n(N)$, получаем, что величины ζ_i из (37) удовлетворяют неравенствам $\zeta_i \leqslant N\mu + o\,(N)$, так что в соответствии с минимаксным представлением (36) имеем $\chi_{n(N)} \leqslant N\mu + o\,(N)$ при $N \to \infty$. С другой стороны, $\zeta_N > N\mu$ при любом выборе d_i, так что $\chi_{n(N)} > N\mu$.

Далее будет более подробно изучен специальный случай $R(N) = 0$ (модель Эрланга, или система с потерями).

Как было показано в [1], в этой ситуации $\beta_{n(N)} \equiv \beta_N$ есть наименьший корень полинома

$$S(x) := \frac{\lambda(N)}{x}\left\{c_{N+1}\left(\frac{x}{\mu}, \frac{\lambda(N)}{\mu}\right) - c_N\left(\frac{x}{\mu}, \frac{\lambda(N)}{\mu}\right)\right\},$$

где

$$c_n(x,a) := \sum_{m=0}^{n} (-1)^m \binom{n}{m}\binom{x}{m}\frac{m!}{a^m}, \qquad n \geqslant 0, \tag{45}$$

— полиномы Шарлье (см., например [5, раздел VI.1]). Поскольку полиномы Шарлье удовлетворяют рекуррентному соотношению

$$ac_{n+1}(x,a) - ac_n(x,a) + xc_n(x-1,a) = 0, \qquad n \geqslant 0,$$

можно записать

$$S(x) = -c_N\left(\frac{x}{\mu} - 1, \frac{\lambda(N)}{\mu}\right),$$

а следовательно,

$$\beta_N = \mu + \mu\xi_{N,1}\left(\frac{\lambda(N)}{\mu}\right), \tag{46}$$

где $\xi_{N,1}(a)$ — минимальный корень полинома Шарлье $c_N(x,a)$. Известно, что $\xi_{N,1}(a) > 0$, так что

$$\beta_N > \mu, \tag{47}$$

однако получение каких-либо явных выражений для $\xi_{N,1}(a)$ в общем случае не представляется возможным. Тем не менее более подробную информацию об асимптотике β_N при $N \to \infty$ получить можно. Случай постоянной $\lambda(N)$ рассмотрен в теореме 11, а теперь будет исследована более интересная ситуация, когда $\lambda(N) = \lambda N$, изучавшаяся, например, в работах [10] и [28].

В случае $\lambda = \mu$ можно найти точное значение β_N. Так, поскольку $c_N(x,a) = c_x(N,a)$ при натуральных x, имеем $c_N(1,N) = c_1(N,N) = 0$.

Однако полиномы Шарлье ортогональны по отношению к мере, состоящей из точечных масс в точках $0, 1, 2, \ldots$, так что из теории ортогональных полиномов (см. [5, гл. 2]) вытекает, что i-й наименьший корень полинома $c_N(x, a)$ больше, чем $i-1$, при $i = 1, 2, \ldots$. В качестве следствия отсюда получаем, что $\xi_{N,1}(N) = 1$, а значит, при всех $N \geqslant 1$

$$\beta_N = 2\mu \quad \text{при} \quad \lambda = \mu. \tag{48}$$

Другой способ доказательства того же равенства (с применением методов предшествующих пунктов) состоит в выборе $d_k = (N-k)(N-k+1)^{-1}$, тогда получаем $\alpha_k = \lambda + \mu + k(N-k)^{-1}(\lambda - \mu)$, а значит, при $\lambda = \mu$ все α_k равны $\lambda + \mu = 2\mu$ и, следовательно, $\beta_N = 2\mu$.

Отметим также, что $c_N(1, a) = c_1(N, a) = 1 - N/a < 0$ при $a < N$. С учетом того, что $c_0(0, a) = 1$, отсюда вытекает оценка $\xi_{N,1}(a) < 1$ при $a < N$ и, значит,

$$\beta_N < 2\mu, \quad \text{если} \quad \lambda < \mu. \tag{49}$$

Используя нашу теорему 7, можно получить следующие асимптотические результаты, согласующиеся с (недоказанными) утверждениями из [10].

Теорема 12. *Пусть* $\lambda(N) = \lambda N$, *причем* $\lambda > 0$, *тогда*

$$\lim_{N\to\infty} \beta_N = \mu, \quad \textit{если} \quad \lambda < \mu,$$
$$\lim_{N\to\infty} \frac{\beta_N}{N} = (\sqrt{\lambda} - \sqrt{\mu})^2, \quad \textit{если} \quad \lambda > \mu.$$

Д о к а з а т е л ь с т в о. Пусть вначале $\lambda < \mu$. Положим $L := [N\lambda/\mu]$ и выберем $d_{N+1} := 0,\ d_N := (N-1)\mu$,

$$d_i := d_{i+1} + (d_{i+1} - d_{i+2})\,\frac{N\lambda}{i\mu}, \qquad i = N-1, N-2, \ldots, L,$$

и $d_i := d_L$ при $i < L$. Понятно, что $d_i > d_{i+1}$, если $i = L, L+1, \ldots, N$, а значит, $d_i > 0$ при $1 \leqslant i \leqslant N$. Подставляя эти d_i в (35), получаем, что $\alpha_i = \mu$ при $i \neq L$. Далее,

$$0 < 1 - \frac{d_{i+1}}{d_i} = \frac{d_{i+1}}{d_i}\left(1 - \frac{d_{i+2}}{d_{i+1}}\right)\frac{N\lambda}{i\mu} < \left(1 - \frac{d_{i+2}}{d_{i+1}}\right)\frac{N\lambda}{i\mu}, \qquad L \leqslant i < N,$$

так что

$$1 - \frac{d_{L+1}}{d_L} < \frac{N\lambda}{L\mu} \cdots \frac{N\lambda}{(N-1)\mu} < \left(\frac{\lambda}{\mu}\right)^{N-L}.$$

В результате получаем

$$\alpha_L = \mu + \left(1 - \frac{d_{L+1}}{d_L}\right)N\lambda < \mu + N\lambda\left(\frac{\lambda}{\mu}\right)^{N-L} = \mu + o\,(1) \quad \text{при} \quad N \to \infty,$$

что, с учетом (47) и минимаксного представления теоремы 7, дает первое утверждение теоремы.

Пусть теперь $\lambda > \mu$. Выбирая $d_i := (\mu/\lambda)^{i/2}$, $i = 1, 2, \ldots, N$, получаем величины α_i из (35) в виде

$$\alpha_i = \begin{cases} (\lambda - \sqrt{\lambda\mu})N - (\sqrt{\lambda\mu} - \mu)i + \sqrt{\lambda\mu}, & 1 \leqslant i < N, \\ \lambda N - (\sqrt{\lambda\mu} - \mu)N + \sqrt{\lambda\mu}, & i = N, \end{cases}$$

так что $\alpha_{N-1} = \min_i\{\alpha_i\}$. По теореме 7 имеем теперь

$$\beta_N \geqslant \alpha_{N-1} = (\sqrt{\lambda} - \sqrt{\mu})^2 N + 2\sqrt{\lambda\mu} - \mu. \tag{50}$$

С другой стороны, полагая $M_1 := [N - \sqrt{N}]$ и $M_2 := [N - \frac{1}{2}\sqrt{N}]$ и выбирая

$$d_i := \left(2\sqrt{\frac{\mu}{\lambda}}\right)^i, \qquad 1 \leqslant i < M_1, \tag{51}$$

и

$$d_{i+1} := \begin{cases} \dfrac{i - M_1 + 3}{i - M_1 + 2}\sqrt{\dfrac{\mu}{\lambda}}\, d_i, & M_1 \leqslant i < M_2, \\ \sqrt{\dfrac{\mu}{\lambda}}\, d_i, & i = M_2, \\ \dfrac{N - i}{N - i + 1}\sqrt{\dfrac{\mu}{\lambda}}\, d_i, & M_2 < i < N, \end{cases} \tag{52}$$

получаем после очевидных преобразований оценки

$$\alpha_i \leqslant (\sqrt{\lambda} - \sqrt{\mu})^2 + o(N) \quad \text{при} \quad N \to \infty, \quad 1 \leqslant i \leqslant N.$$

А тогда, в соответствии с минимаксным представлением теоремы 7, получаем

$$\beta_N \leqslant (\sqrt{\lambda} - \sqrt{\mu})^2 N + o(N) \quad \text{при} \quad N \to \infty, \tag{53}$$

что вместе с (50) дает второе утверждение теоремы.

З а м е ч а н и е 1. Используя совершенно иную технику, в [19, теорема 3] показано, что $\xi_{N,1}(aN) > (\sqrt{a} - 1)^2 N + c\sqrt[3]{N}$ при $a > 1$ для некоторых положительных постоянных c, так что (50) может быть заменено на $\beta_N > (\sqrt{\lambda} - \sqrt{\mu})^2 N + c\mu\sqrt[3]{N}$.

З а м е ч а н и е 2. Другое доказательство (53) можно получить, используя схему рассуждений леммы 36 из [21, гл. 17] и соотношение

$$ac_n(x + 1, a) + (n - a - x)c_n(x, a) + xc_n(x - 1, a) = 0, \qquad n \geqslant 0.$$

(Этот подход был предложен И. Красиковым.)

Можно использовать $\mathbf{d}$ из доказательства предыдущей теоремы, чтобы получить верхние оценки для $\|\mathbf{p}(t) - \boldsymbol{\pi}\|$ с помощью оценки (38). А именно, получаем:

$$\|\mathbf{p}(t) - \boldsymbol{\pi}\| \leqslant \begin{cases} C_1 e^{-\{N(\sqrt{\lambda} - \sqrt{\mu})^2 + 2\sqrt{\lambda\mu} - \mu\}t} & \text{при} \quad \lambda > \mu, \\ C_2 e^{-\mu t} & \text{при} \quad \lambda < \mu, \end{cases} \qquad t \geqslant 0, \tag{54}$$

где постоянные C_1 и C_2 зависят от N. Эти оценки согласуются с неравенствами из [10, утверждения 6 и 10].

Рассмотрим, наконец, асимптотическое поведение χ_N при $N \to \infty$ в условиях предыдущей теоремы.

Теорема 13. *Пусть* $\lambda(N) = \lambda N$ *при* $\lambda > 0$, *тогда*

$$\lim_{N\to\infty} \frac{\chi_N}{N} = (\sqrt{\lambda} + \sqrt{\mu})^2.$$

Доказательство. Выбирая $d_i := (\mu/\lambda)^{i/2}$, $i = 1, \ldots, N$, получаем величины ζ_i из (37) в виде:

$$\zeta_i = \begin{cases} (\lambda + \sqrt{\lambda\mu})N + (\sqrt{\lambda\mu} + \mu)i - \sqrt{\lambda\mu}, & 1 \leqslant i < N, \\ \lambda N + (\sqrt{\lambda\mu} + \mu)N - \sqrt{\lambda\mu}, & i = N, \end{cases}$$

так что при достаточно больших N имеем $\zeta_{N-1} = \max_i\{\zeta_i\}$. Тогда из (36) следует, что

$$\chi_N \leqslant \zeta_{N-1} = (\sqrt{\lambda} + \sqrt{\mu})^2 N - 2\sqrt{\lambda\mu} - \mu. \tag{55}$$

С другой стороны, выбирая d_i так же, как в (51) и (52), получаем оценки

$$\zeta_i \geqslant (\sqrt{\lambda} + \sqrt{\mu})^2 + o(N) \quad \text{при} \quad N \to \infty, \quad 1 \leqslant i \leqslant N,$$

а значит, с учетом минимаксного представления (36), имеем

$$\chi_N \geqslant (\sqrt{\lambda} + \sqrt{\mu})^2 N + o\,(N) \quad \text{при} \quad N \to \infty. \tag{56}$$

Наконец, применяя (55) и (56), получаем утверждение теоремы.

Авторы благодарны Илье Красикову за полезные обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. *Blanc J. P. C., van Doorn E. A.* Relaxation times for queueing systems. — Mathematics and Computer Science (Amsterdam, 1983). Ed. by J. W. de Bakker, M. Hazewinkel, and J. K. Lenstra. Amsterdam: North-Holland, 1986, p. 139–162. (CWI Monogr., v. 1.)
2. *Chen M.-F.* Estimation of spectral gap for Markov chains. — Acta Math. Sinica (N.S.), 1996, v. 12, № 4, p. 337–360.
3. *Chen M.-F.* Variational formulas and approximation theorems for the first eigenvalue in dimension one. — Sci. China Ser. A, 2001, v. 44, № 4, p. 409–418.
4. *Chen M.-F.* Eigenvalues, Inequalities, and Ergodic Theory. London: Springer-Verlag, 2005, 228 p.
5. *Chihara T. S.* An Introduction to Orthogonal Polynomials. New York: Gordon and Breach, 1978, 249 p.
6. *Dahlquist G.* Stability and Error Bounds in the Numerical Integration of Ordinary Differential Equations. Inaugural dissertation (University of Stockholm). Uppsala: Almqvist & Wiksells Boktryckeri AB, 1958, 87 p. Reprinted in: Transactions of the Royal Institute of Technology, № 130. Stockholm, 1959.

7. *van Doorn E. A.* Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process. — Adv. Appl. Probab., 1985, v. 17, № 3, p. 514–530.
8. *van Doorn E. A.* Representations for the rate of convergence of birth-death processes. — Theory Probab. Math. Statist., 2002, v. 65, p. 37–43.
9. *van Doorn E. A., van Foreest N. D., Zeifman A. I.* Representations for the extreme zeros of orthogonal polynomials. — J. Comput. Appl. Math. (to appear); http://eprints.eemcs.utwente.nl/10994/
10. *Fricker C., Robert P., Tibi D.* On the rates of convergence of Erlang's model. — J. Appl. Probab., 1999, v. 36, № 4, p. 1167–1184.
11. *Гнеденко Б. В., Макаров И. П.* Свойства решений задачи с потерями в случае периодических интенсивностей. — Дифференц. уравнения, 1971, т. 7, с. 1696–1698.
12. *Granovsky B. L., Zeifman A. I.* The decay function of nonhomogeneous birth-death processes, with application to mean-field models. — Stochastic Process. Appl., 1997, v. 72, № 1, p. 105–120.
13. *Granovsky B. L., Zeifman A. I.* The N-limit of spectral gap of a class of birth-death Markov chains. — Appl. Stoch. Models Bus. Ind., 2000, v. 16, № 4, p. 235–248.
14. *Granovsky B. L., Zeifman A. I.* Nonstationary queues: estimation of the rate of convergence. — Queueing Syst., 2004, v. 46, № 3, p. 363–388.
15. *Грановский Б. Л., Зейфман А. И.* О нижней границе спектра для некоторых моделей среднего поля. — Теория вероятн. и ее примен., 2004, т. 49, в. 1, с. 164–171.
16. *Karlin S., McGregor J. L.* The differential equations of birth-and-death processes, and the Stieltjes moment problem. — Trans. Amer. Math. Soc., 1957, v. 85, p. 489–546.
17. *Karlin S., McGregor J. L.* Ehrenfest urn models. — J. Appl. Probab., 1965, v. 2, p. 352–376.
18. *Kijima M.* Evaluation of the decay parameter for some specialized birth-death processes. — J. Appl. Probab., 1992, v. 29, № 4, p. 781–791.
19. *Krasikov I.* Bounds for zeros of the Charlier polynomials. — Methods Appl. Anal., 2002, v. 9, № 4, p. 599–610.
20. *Лозинский С. М.* Оценка погрешности численного интегрирования обыкновенного дифференциального уравнения, I. — Изв. вузов. Матем., 1958, в. 5, с. 52–90; исправл.: 1959, в. 5, с. 222.
21. *MacWilliams F. J., Sloane N. J. A.* The Theory of Error-Correcting Codes. Part II. Amsterdam: North-Holland, 1977, 392 p.
22. *Meyer C. D.* Matrix Analysis and Applied Linear Algebra. Philadelphia: SIAM, 2001, 718 p. (Updates available on http://www.matrixanalysis.com).
23. *Рудин У.* Функциональный анализ. М.: Мир, 1975, 443 с.
24. *Söderlind G.* The logarithmic norm. History and modern theory. — BIT. Numerical Mathematics, 2006, v. 46, № 3, p. 631–652.
25. *Ström T.* On logarithmic norms. — SIAM J. Numer. Anal., 1975, v. 12, № 5, p. 741–753.
26. *Takács L.* Introduction to the Theory of Queues. New York: Oxford Univ. Press, 1962, 268 p.
27. *Varga R. S.* Matrix Iterative Analysis. Berlin: Springer-Verlag, 2000, 358 p.
28. *Voit M.* A note on the rate of convergence to equilibrium for Erlang's model in the subcritical case. — J. Appl. Probab., 2000, v. 37, № 3, p. 918–923.
29. *Zeifman A. I.* Some estimates of the rate of convergence for birth and death processes. — J. Appl. Probab., 1991, v. 28, № 2, p. 268–277.
30. *Zeifman A. I.* On the estimation of probabilities for birth and death processes. — J. Appl. Probab., 1995, v. 32, № 3, p. 623–634.
31. *Zeifman A. I.* Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. — Stochastic Process. Appl., 1995, v. 59, № 1, p. 157–173.
32. *Зейфман А. И., Бенинг В. Е., Соколов И. А.* Марковские цепи и модели с непрерывным временем. М.: ЭЛЕКС-КМ, 2008.

Поступила в редакцию
17.IX.2008