PDMI Preprint 11/2022

Вероятностные свойства избыточных систем счисления

Н. В. Дуров[1]

[1] Санкт-Петербургское отделение Математического института им. В. А. Стеклова Российской Академии Наук

email: douroff@pdmi.ras.ru

Аннотация. Данная работа посвящена изучению некоторых вероятностных свойств эффективных систем счисления, примеры которых были построены ранее в [AB4]. В частности, рассматривается вопрос об естественном распределении вероятностей на цифрах таких систем, и о подсчёте количества различных n-значных чисел в таких системах. Оказывается, что конечный автомат, проверяющий равенство в эффективной системе счисления, может быть естественным образом преобразован в некоторую цепь Маркова, свойства которой тесно связаны с упомянутыми выше вопросами. Мы изучаем спектральный радиус и спектр матрицы переходов этой цепи Маркова, и даём полный ответ для некоторого интересного класса эффективных систем счислений. При этом возникает интересный новый класс многочленов, похожих на многочлены Эрмита или на многочлены Кравчука.

Ключевые слова: Абсолютная геометрия, эффективная система счисления, цепь Маркова, матрица переходов цепи Маркова, многочлены Эрмита, многочлены Кравчука.

ПРЕПРИНТЫ ПОМИ РАН

Учредитель: Федеральное государственное бюджетное учреждение науки
Санкт-Петербургское отделение Математического института
им. В. А. Стеклова Российской академии наук

Свидетельство о регистрации средства массовой информации:
ЭЛ № ФС 77-33560 от 16 октября 2008 г.
Выдано Федеральной службой по надзору
в сфере связи и массовых коммуникаций

Контактные данные:

191023, г. Санкт-Петербург, наб. реки Фонтанки, дом 27

телефоны: (812) 312-40-58; (812) 571-57-54

e-mail: admin@pdmi.ras.ru

http://www.pdmi.ras.ru/preprint/

Заведующая информационно-издательским сектором В. Н. Симонова

5 Вероятностные свойства избыточных систем счисления

5.1 Информационный парадокс

Прежде, чем двигаться дальше, попробуем ответить на вопрос: сколько информации содержится в одной цифре числа, записанного в избыточной двоичной системе с цифрами $D \in \{-1, 0, 1\}$ из [AB4],**4.1**?

5.1.1. (Каждая цифра уменьшает диапазон значений вдвое, и потому содержит один бит информации.) Пусть некто сообщает неизвестное нам число $\xi \in I := [-1, 1]$ по цифрам его избыточной двоичной записи $\xi = (0.x_0x_1\ldots)_2$. Будем считать, что изначально нам ничего не известно о ξ, так что естественно предполагать его равнораспределённым на этом отрезке. Далее, после того, как мы узнаём первую цифру x_0, множество возможных значений ξ уменьшается до одного из отрезков вдвое меньшей длины $I_l := [-1, 0]$, $I_m := [-1/2, 1/2]$ или $I_r := [0, 1]$, которые мы будем называть «левым», «средним» и «правым подотрезком», соответственно. Таким образом, цифра $x_0 \in D$ осуществляет выбор одного из этих трёх отрезков. После этого цифра $x_1 \in D$ снова уменьшает выбранный отрезок вдвое, заменяя его на его левый, средний или правый подотрезок, и т.д.

Поскольку очередная сообщаемая нам цифра уменьшает пространство возможных значений ξ ровно вдвое (т.е. мера Лебега множества возможных значений уменьшается ровно в два раза), это означает, что она даёт нам ровно один бит информации о вещественном числе ξ.

5.1.2. (Парадокс: каждая цифра выбирается из трёх вариантов, и потому несёт $\log_2 3$ битов информации.) С другой стороны, каждая сообщаемая нам цифра выбирается из трёхэлементного множества D, и потому её указание требует $\log_2 3$ битов информации. Иначе говоря, мы должны считать, что каждая цифра несёт $\log_2 3 \approx 1.853 > 1$ битов информации, что противоречит предыдущему пункту.

5.1.3. (Другой вариант парадокса с неравновероятным выбором цифр.) Мы видели в [AB4],**4.1.32**, что естественное распределение вероятностей на наборе цифр D — это вовсе не равномерное распределение, а $P(0) = 1/2$, $P(1) = P(\bar{1}) = 1/4$. Например, если отправитель получил отправляемую нам избыточную двоичную запись числа ξ в результате

каких-либо арифметических действий, скажем, суммирования большого количества слагаемых, то рассуждения из [AB4],**4.1.32** показывают, что частоты цифр будут примерно такими.

Таким образом, нам естественно ожидать именно такое распределение частот или вероятностей символов, принимаемых нами по каналу. Это означает, что один символ несёт не $\log_2 3$ битов информации, как это было бы в случае равномерного распределения, а энтропию указанного выше распределения

$$H(P) = \sum_{x\in D} -P(x)\log_2 P(x) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = \frac{3}{2} \qquad (5.1.3.1)$$

т.е. полтора бита информации на цифру. Это все равно больше единицы из **5.1.1**.

5.1.4. (Частичное решение парадокса: неравенство обработки данных.) Конечно же, парадокс возникает из-за того, что мы смешиваем два понятия: вещественное число $\xi \in [-1,1]$ и его конкретную запись $\boldsymbol{x} = x_0x_1x_2\ldots \in D^\omega$. Эти две величины связаны отображением $w(\boldsymbol{x}) = \xi$, однако w сюръективно, но не биективно, потому что система счисления избыточна. Поэтому конкретная запись числа действительно содержит больше информации, чем само число. Однако после получения вещественного числа ξ мы можем «забыть» его конкретную запись и работать «с самим числом» (для чего мы можем преобразовать полученную запись в какую-нибудь другую, например, в избыточную шестнадцатеричную систему с цифрами $-8\ldots 8$, что существенно снизит избыточность). В этот момент мы заменяем ξ на значение некоторой функции $w(\xi)$, а известно, что при таком переходе количество информации (энтропия случайной величины) не увеличивается (это частный случай неравенства обработки данных – data processing inequality). Иначе говоря, применение w неизбежно теряет часть информации, и именно это мы и наблюдаем.

5.1.5. (Избыточная двоичная система счисления не соответствует мере Лебега на отрезке $[-1,1]$, а соответствует мере с плотностью $1-|x|$.) Для формального применения неравенства обработки данных нам было бы полезно знать, что $w(\boldsymbol{x})$ равномерно распределено на отрезке $[-1,1]$, в ситуации, когда цифры $\boldsymbol{x} = x_0x_1x_2\ldots \in D^\omega$ распределены независимо, например, с вероятностями $P(0) = 1/2$, $P(1) = P(\bar{1}) = 1/4$ как выше. Иначе говоря, мы бы хотели, чтобы прямой образ относительно

$w : D^\omega \to [-1,1]$ меры-прямого произведения на D^ω равнялся бы мере Лебега на $[-1,1]$. Однако в действительности аналогичное утверждение верно только для обычной (не избыточной) двоичной системы, а для избыточной двоичной системы образ меры оказывается равен не мере Лебега на $[-1,1]$, а мере $\big(1-|x|\big)dx$ с «треугольной» плотностью. Для того, чтобы понять это, можно записать случайную цифру $x_i \in \{-1,0,1\}$ в виде $x_i = (\varepsilon_i^{(1)} + \varepsilon_i^{(2)})/2$, где $\varepsilon_i^{(1)}$ и $\varepsilon_i^{(2)}$ — независимые случайные величины, принимающие значения ± 1 с вероятностью $1/2$. Тогда

$$w(x) = \sum_{i=0}^{\infty} x_i 2^{-i-1} = \frac{1}{2}\Big(\sum_{i=0}^{\infty} \varepsilon_i^{(1)} 2^{-i-1} + \sum_{i=0}^{\infty} \varepsilon_i^{(2)} 2^{-i-1}\Big)$$

оказывается полусуммой двух независимых случайных величин, равномерно распределённых на $[-1,1]$, а значит, плотность распределения случайной величины $w(x)$ равна свёртке $\chi_{[-1/2,1/2]} * \chi_{[-1/2,1/2]}$, где $\chi_A(x) = [x \in A]$ обозначает характеристическую функцию множества $A \in \mathbb{R}$. Эта свёртка и есть $\big(1-|x|\big)dx$ на отрезке $[-1,1]$.

5.1.6. (Образ меры относительно $w : D^\omega \to [-1,1]$ для других вероятностей цифр.) Мы можем задаться вопросом, каков будет образ меры-произведения на D^ω относительно w для других распределений вероятностей цифр из D. В общем случае ответ получается следующий:

- Если $P(0) \neq 1/2$, т.е. если чётные и нечётные цифры не равновероятны, то мера концентрируется возле чисел, в обычной двоичной записи которых нули преобладают над единицами или наоборот, и в итоге получается мера, не обладающая плотностью относительно меры Лебега.

- Если $P(0) = 1/2$, $P(1) = \alpha/2$, $P(\bar{1}) = \beta/2$, $0 \le \alpha \le 1$, $\beta = 1-\alpha$, то получается мера с плотностью $f_\alpha(t)$, которую можно найти из функционального уравнения

  $$f_\alpha(t) = f_\alpha(2t) + \alpha f_\alpha(2t-1) + (1-\alpha) f_\alpha(2t+1) \quad \forall t \in \mathbb{R} \quad (5.1.6.1)$$

  с дополнительным условием $f_\alpha(t) = 0$ при $t \notin [-1,1]$, выражающими требование, чтобы случайная величина ξ с такой плотностью распределения и $(\xi+\eta)/2$, где η — независимая от ξ случайная цифра из D с заданным распределением вероятностей, имели бы одинаковое распределение.

5.2 Количество n-значных чисел

Информационный парадокс из предыдущего пункта допускает следующую чисто комбинаторную переформулировку: «Сколько существует n-значных чисел в избыточной двоичной системе?» Отнесёмся к этому вопросу как можно более серьёзно, поскольку попытки найти удовлетворительный ответ на него приводят к довольно глубоким и неожиданным результатам.

5.2.1. (Количество n-значных чисел в избыточной двоичной системе.) Итак, сколько же существует n-значных двоичных чисел

$$x = (x_0 x_1 \dots x_{n-1})_2 = \sum_{i=0}^{n-1} x_i 2^{n-1-i} \quad \text{где } x_i \in \{-1, 0, 1\} \tag{5.2.1.1}$$

в избыточной двоичной системе? Ясно, что есть два разных ответа:

- Это слова $x_0 x_1 \dots x_{n-1}$ длины n в трёхсимвольном алфавите $D = \{-1, 0, 1\}$, а значит, их 3^n штук. Более изощрённый вариант подсчёта — $2^{3n/2} \approx 2.82^n$ штук, если не считать цифры из D равновероятными, см. **5.1.3**.
- Поскольку $\{x = \sum_{i=0}^{n-1} x_i 2^{n-1-i} : x_i \in D\}$ — это в точности множество всех целых чисел от $1 - 2^n$ до $2^n - 1$, их $2^{n+1} - 1$ штука.

5.2.2. (Отношение эквивалентности $\equiv$ на D^n.) Ясно, что $2^{n+1} - 1 < 3^n$ при $n \geq 2$, и что разница возникает из-за того, что числа соответствуют не словам из D^n, а классам эквивалентности слов относительно некоторого отношения эквивалентности $\equiv$, которое мы описали достаточно явно в [AB4],**4.2.5**. Можем ли мы воспользоваться тем, что $\equiv$ допускает эффективное описание с помощью конечного автомата из [AB4],**4.4**, чтобы вычислить интересующее нас количество классов эквивалентности чисто комбинаторным способом, не зная, что это в действительности избыточные двоичные записи целых чисел?

5.2.3. (Нам нужно только $\equiv$, а про $\oplus$ и $\ominus$ пока можно забыть.) Отметим, что для наших нынешних целей нам нужен только конечный автомат, задающий отношение эквивалентности $\equiv$, например, на конечных словах D^n, а про конечные автоматы, задающие $\oplus$ и $\ominus$, мы можем забыть. В каком-то смысле забывание сложения соответствует сужению скаляров с $\mathbb{Q}$ до $\mathbb{F}_1$ — или скорее до $\mathbb{F}_1[t, t^{-1}]$, поскольку у нас остаётся операция

сдвига бесконечного слова вправо или влево, соответствующая умножению на степени переменной t.

5.2.4. (Подсчёт количества пар эквивалентных слов с помощью степеней матрицы.) Теперь мы можем легко подсчитать количество пар $(x, y) = (x_0 \dots x_{n-1}, y_0 \dots y_{n-1})$ эквивалентных слов длины n, поскольку такие пары соответствуют путям длины n из начального состояния C в себя в конечном автомате [AB4],**4.4.3**, распознающем эквивалентность слов. Таким образом, надо взять матрицу смежности графа переходов этого автомата (с учётом кратности рёбер)

$$A = \begin{pmatrix} 1 & & & & \\ 3 & 2 & 1 & & \\ 1 & 2 & 3 & 2 & 1 \\ & & 1 & 2 & 3 \\ & & & & 1 \end{pmatrix} \tag{5.2.4.1}$$

и посмотреть на центральный элемент её степени A^n. Например, таким образом можно установить, что существует 2423293 пары слов $(x, y) \in D^{10} \times D^{10}$, такие, что $x \equiv y$.

5.2.5. (Приблизительный подсчёт количества классов эквивалентности.) Однако нам нужно узнать количество классов эквивалентности $D^n/\equiv$, а не количество эквивалентных пар. Если бы речь шла о факторгруппах конечных групп, то все классы эквивалентности состояли бы из одинакового количества элементов, и мы получили бы

$$\left|D^n/\equiv\right| \approx \frac{|D^n|^2}{\left|\{(x,y) : x \equiv y\}\right|} \tag{5.2.5.1}$$

В нашем случае мы сделаем вид, что эта формула приблизительно верна. Можно предложить неформальное обоснование в духе доказательств теорем Шеннона. Предположим, большинство элементов D^n «типичны», и классы эквивалентности типичных элементов состоят приблизительно из одинакового количества элементов M. Тогда у нас есть примерно $|D^n|/M$ классов эквивалентности типичных элементов, и примерно $|D^n|^2/M$ пар эквивалентных типичных элементов. Отсюда снова получаем (5.2.5.1), при условии, что типичных элементов много.

Например, в нашем случае мы могли бы заключить, что существует примерно $3^{20}/2423293 \approx 1439$ классов эквивалентности в $D^{10}/\equiv$. Это уже не так далеко от правильного значения 2047, как $3^{10} = 59049$.

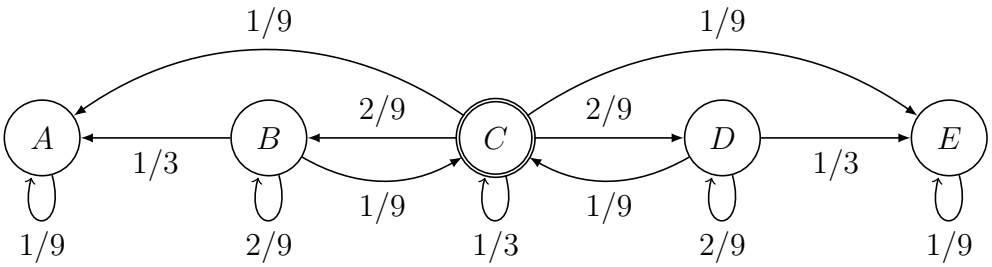
5.2.6. (Поведение приблизительной формулы при $n \to +\infty$.) При $n \to +\infty$ значение центрального элемента матрицы A^n ведёт себя приблизительно как $c\lambda^n$, где $\lambda = (5+\sqrt{17})/2 \approx 4.56$ — наибольшее по модулю собственное число матрицы A, а $c = (1+1/\sqrt{17})/2 \approx 0.621$ — произведение подходящих элементов матрицы собственных векторов матрицы A и обратной к ней. Поэтому приведённый выше метод даёт в качестве количества n-значных чисел приблизительно $c^{-1}(9/\lambda)^n \approx 1.61 \cdot 1.973^n$. Это уже гораздо ближе к правильному ответу $2^{n+1}-1$, но ещё не совсем он.

5.2.7. (Вероятностная интерпретация предыдущего вычисления.) Заметим, что в качестве приблизительного значения количества n-значных чисел у нас получилось число, обратное к центральному элементу матрицы $(A/9)^n$:

$$\frac{A}{9} = \begin{pmatrix} 1/9 & & & & \\ 1/3 & 2/9 & 1/9 & & \\ 1/9 & 2/9 & 1/3 & 2/9 & 1/9 \\ & & 1/9 & 2/9 & 1/3 \\ & & & & 1/9 \end{pmatrix} \tag{5.2.7.1}$$

Эта матрица допускает следующую вероятностную интерпретацию. Пусть x и y — два случайных слова длины n, все символы которых независимо друг от друга принимают все три возможных значения из $D = \{-1, 0, 1\}$ с одинаковой вероятностью $1/3$. С какой вероятностью $x \equiv y$?

Будем проверять, верно ли, что $x \equiv y$, последовательно применяя переходы с метками (x_i, y_i) в нашем конечном автомате G. Ясно, что очередной переход выбирается случайно и независимо от предыдущего, причём вероятность каждой пары xy одинакова и равна $1/9$. В итоге конечный автомат [AB4],**4.4.3** становится цепью Маркова со следующими вероятностями переходов:

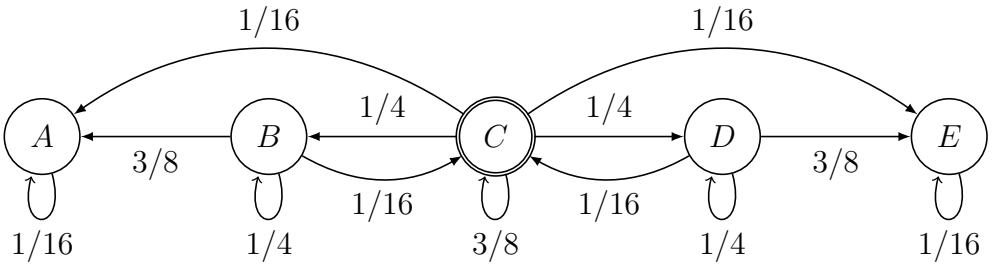


Матрица $A/9$ из (5.2.7.1) — это матрица переходов этой цепи Маркова. Центральный элемент матрицы $(A/9)^n$ — это вероятность p того, что эта цепь Маркова, начав работу с состояния C, окажется снова в состоянии C через n шагов. Иначе говоря, p — это вероятность того, что два случайно выбранных слова длины n эквивалентны, т.е. того, что два n-значных числа равны. Поэтому $1/p$ — оценка количества различных n-значных чисел.

5.2.8. (Это не совсем цепь Маркова.) Здесь есть небольшая тонкость, связанная с тем, что сумма вероятностей переходов из каждого состояния нашей «цепи Маркова» в действительности не всегда равна единице, а может быть меньше, из-за того, что часть переходов запрещена. Мы могли бы изготовить настоящую цепь Маркова, добавив специальное «ошибочное состояние» и сделав так, чтобы все запрещённые переходы вели в него. Однако в этом нет особого смысла, кроме того, что таким образом становится выполнено формальное определение цепи Маркова. Например, вероятность перехода из состояния C в себя за n шагов не изменяется. Зато матрица переходов $A/9$ приобретает лишнюю строку и столбец и становится менее красивой.

Отметим, что именно отсутствие части переходов (или, если угодно, неэргодичность нашей цепи Маркова) и делает ответ в нашей задаче нетривиальным. Иначе наибольшим собственным значением матрицы переходов была бы единица, и мы получали бы ответы порядка $O(1)$ при всех n.

5.2.9. (Оценка с помощью естественного распределения вероятностей цифр.) Заметим, что предыдущее вероятностное рассуждение с независимым выбором цифр x и y и вычислением вероятности того, что $x \equiv y$, с помощью цепи Маркова, работает и в том случае, если цифры не равновероятны, а, например, выбираются согласно более естественному распределению $P(0) = 1/2$, $P(1) = P(\bar{1}) = 1/4$ из [AB4],**4.1.32**. В этом случае получаем другую цепь Маркова



с матрицей переходов

$$P = \begin{pmatrix} 1/16 & & & & \\ 3/8 & 1/4 & 1/16 & & \\ 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \\ & & 1/16 & 1/4 & 3/8 \\ & & & & 1/16 \end{pmatrix} = \frac{1}{16} \begin{pmatrix} 1 & & & & \\ 6 & 4 & 1 & & \\ 1 & 4 & 6 & 4 & 1 \\ & & 1 & 4 & 6 \\ & & & & 1 \end{pmatrix} \qquad (5.2.9.1)$$

Собственные числа этой матрицы (неслучайно) равны 1/2, 1/4, 1/8, 1/16 и 1/16, и в итоге вероятность p того, что $x \equiv y$ для слов длины n, оказывается равна $(2/3) \cdot 2^{-n} + O(4^{-n})$, а оценка на количество n-значных чисел — $1.5 \cdot 2^n + O(1)$, что уже отличается от правильного ответа $2^{n+1} - 1$ только постоянным множителем.

5.2.10. (Вероятностный метод всегда даёт оценку снизу.) Отметим, что применяемый нами вероятностный метод всегда даёт оценку снизу на число классов эквивалентности. В самом деле, пусть существует N классов эквивалентности слов длины n, и пусть вероятность того, что случайно выбранное число (с заданным распределением вероятностей цифр) попадает в i-ый класс эквивалентности, равно p_i. Таким образом, $\sum_{i=1}^{N} p_i = 1$, и при этом вероятность того, что два случайно выбранных слова эквивалентны, равна $p = \sum_{i=1}^{N} p_i^2$. Отсюда по неравенству о среднем арфиметическом и среднем квадратическом получаем $p/N \geq (1/N)^2$, т.е. $p \geq 1/N$ и $N \geq p^{-1}$.

5.2.11. (Другие распределения вероятностей цифр не помогают улучшить оценку.) Мы можем рассмотреть другие распределения вероятностей цифр, и посмотреть, какие оценки на количество n-значных чисел будут получаться. Систематический способ это сделать — рассмотреть все такие распределения $P = (p_{-1}, p_0, p_1)$, где $p_{-1}, p_0, p_1 \geq 0$, $p_{-1} + p_0 + p_1 =$

1, построить для каждого из них соответвующую цепь Маркова с матрицей переходов

$$T(P) := \begin{pmatrix} c & & & & \\ a & b & c & & \\ c & b & a & b & c \\ & & c & b & a \\ & & & & c \end{pmatrix} \tag{5.2.11.1}$$

где $a = p_0^2 + p_1^2 + p_{-1}^2$, $b = p_0(p_1 + p_{-1})$, $c = p_1 p_{-1}$, и затем найти те из них, что минимизируют спектральный радиус $\lambda = \rho(T(P))$ (наибольшее по модулю собственное число, которое обязательно будет положительным), поскольку в конечном итоге мы получим оценку снизу вида $c^{-1}\lambda^{-n}$ на количество n-значных чисел, а значит, надо, чтобы λ было минимальным. В нашем случае минимальный спектральный радиус оказывается $1/2$, однако достигается он не в одной точке, а на отрезке $\{p_0 = 1/2, p_{-1} = 1/2 - p_1\}$. Для каждой из точек этого отрезка мы можем вычислить коэффициент $c = c(P)$, который получается произведением подходящих элементов матрицы собственных векторов $T(P)$ и обратной к ней. В нашем случае оказывается, что $c(P) = 1/(1 + 4p_1 - 8p_1^2)$ при $p_0 = 1/2$, $p_{-1} = 1/2 - p_1$, что даёт оценку снизу вида $(1 + 4p_1 - 8p_1^2) \cdot 2^n$ на количество n-значных чисел. Естественно попробовать минимизировать $c(P)$ на этом отрезке, т.е. максимизировать $1 + 4p_1 - 8p_1^2$ при $0 \le p_1 \le 1/2$. В данном случае максимум достигается в середине отрезка при $p_1 = 1/4$, что соответствует нашему «естественному распределению» $P(0) = 1/2$, $P(1) = P(\bar{1}) = 1/4$, т.е. соответствующая оценка $1.5 \cdot 2^n$ в действительности является наилучшей (при $n \to +\infty$) среди всех оценок вида $c^{-1} \cdot \lambda^{-n}$. Иначе говоря, оценку, полученную в **5.2.9** с помощью естественного распределения, уже не получается улучшить тем же методом.

5.2.12. (Естественное распределение вероятностей цифр полностью определяется конечным автоматом.) Отметим, что естественное распределение вероятностей цифр из D, ранее полученное в [AB4],**4.1.32** из других соображений, оказалось полностью определено чисто комбинаторной структурой конечного автомата G, проверяющего конечные или бесконечные слова на эквивалентность, поскольку оно получается последовательной минимизацией сначала спектрального радиуса матрицы переходов $T(P)$ на (компактном) симплексе вероятностных мер на D, а затем минимизацией коэффициента $c(P)$. В этом смысле естественное распре-

деление вероятностей цифр действительно естественно.

5.2.13. (Предельное распределение вероятностей состояний конечного автомата.) Отметим, что комбинаторная структура конечного автомата проверки эквивалентности слов определяет не только естественное распределение вероятностей цифр, но и предельное распределение вероятностей на состояниях самого конечного автомата. Это предельное распределение вероятностей на состояниях марковской цепи **5.2.9** (после большого числа шагов, при условии, что работа автомата не завершилась ранее), оно задаётся собственным ковектором матрицы переходов P из (5.2.9.1), соответствующим максимальному собственному числу $\lambda = 1/2$:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} P = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{5.2.13.1}$$

Иначе говоря, *в пределе все состояния автомата равновероятны:* $P(A) = P(B) = P(C) = P(D) = P(E) = 1/5$.

5.2.14. (Предельное распределение вероятностей состояний для «развёрнутого» автомата.) Интересно, что при использовании автомата в обратную сторону — то есть для проверки равенства 2-адических чисел — матрица переходов P заменяется на транспонированную P^T, и предельное распределение вероятностей состояний конечного автомата оказывается совершенно другим: $P(A) = P(E) = 0$, $P(B) = P(D) = 1/6$, $P(C) = 2/3$. Поэтому в общем случае не стоит ожидать, что все состояния конечного автомата проверки эквивалентности в пределе равновероятны.

5.2.15. (Пути улучшения оценки количества n-значных чисел.) Вернёмся к проблеме подсчёта количества n-значных чисел. Как можно улучшить оценку $1.5 \cdot 2^n$? Можно предположить, что наша оценка занижена из-за того, что слишком много слов длины n, пусть даже с естественным распределением частот цифр, оказались «нетипичны», и в их классах эквивалентности слишком много или слишком мало элементов. Нетипичность, скорее всего, связана с большим количеством «экстремальных» цифр 1 и $\bar{1}$, поскольку именно при сравнении чисел с большим количеством таких цифр автомат оказывается в крайних состояниях вроде A и E, в которых нарушается эргодичность. Если бы мы могли понизить частоту экстремальных цифр, это увеличило бы долю типичных чисел. Однако мы знаем, что в данном случае мы не можем отойти от естественного распределения $P(1) = P(\bar{1}) = 1/4$, и примерно половина всех цифр всегда будут экстремальными.

Напрашивается следующий путь решения возникающей проблемы: расширить множество допустимых цифр до $\tilde{D} = \{-2, -1, 0, 1, 2\}$, но сделать новые экстремальные цифры -2 и 2 гораздо более редкими, чем раньше. Тогда доля типичных n-значных чисел вырастет, и оценка станет более точной.

5.2.16. (Случай $\tilde{D} = \{-2, -1, 0, 1, 2\}$.) Итак, рассмотрим избыточную двоичную систему с пятью цифрами $\tilde{D} = \{-2, -1, 0, 1, 2\}$. Теперь n-значные числа — это целые числа от $-2^{n+1} + 1$ до $2^{n+1} - 1$, и их всего $2^{n+2} - 1$ штук. Слов длины n теперь 5^n, но мы уже знаем, как с этим бороться. В качестве естественного распределения частот цифр из $\tilde{D}$ сразу возьмём биномиальное распределение $P(2) = P(\bar{2}) = 1/16$, $P(1) = P(\bar{1}) = 1/4$, $P(0) = 3/8$. Такой выбор можно неформально обосновать, например, тем, что случайное n-значное число с цифрами из $\tilde{D}$ можно получить, сложив без переносов два случайных n-значных числа с цифрами из D. Более сложный способ найти правильное распределение — рассмотреть все вероятностные распределения и минимизировать спектральный радиус матрицы переходов марковской цепи, а затем и коэффициент c, как это было объяснено в **5.2.12**. Мы, однако, поверим, что мы угадали правильный ответ, и сразу выпишем матрицу переходов марковской цепи именно для такого выбора вероятностей цифр:

$$\tilde{P} = \frac{1}{256}\begin{pmatrix}
1 & & & & & & & & \\
28 & 8 & 1 & & & & & & \\
70 & 56 & 28 & 8 & 1 & & & & \\
28 & 56 & 70 & 56 & 28 & 8 & 1 & & \\
1 & 8 & 28 & 56 & 70 & 56 & 28 & 8 & 1 \\
 & & 1 & 8 & 28 & 56 & 70 & 56 & 28 \\
 & & & & 1 & 8 & 28 & 56 & 70 \\
 & & & & & & 1 & 8 & 28 \\
 & & & & & & & & 1
\end{pmatrix} \tag{5.2.16.1}$$

Здесь состояния автомата соответствуют $s = -4, -3, \ldots, 4$ в алгоритме, аналогичном [AB4],**4.4.1**. Элементы матрицы $\tilde{P}$ — это, конечно же, биномиальные коэффициенты:

$$\tilde{P}_{st} = \begin{cases} 2^{-8}\binom{8}{2s-t+4} & \text{при } |2s - t| \le 4 \\ 0 & \text{иначе} \end{cases} \tag{5.2.16.2}$$

где индексы пробегают значения $-4 \le s,t \le 4$. Собственные значения матрицы $\tilde{P}$ оказываются равны $1/2$, $1/4$, $1/8$, …, $1/128$, $1/256$, $1/256$. Максимальное собственное значение $\lambda = 1/2$, а коэффициент $c = 151/315$. В итоге получаем оценку[1] $c^{-1}\lambda^{-n} = \frac{315}{151} \cdot 2^n \approx 2.086 \cdot 2^n$. Это почти вдвое меньше правильного ответа $4 \cdot 2^n - 1$.

5.2.17. (Цифры $\{-3,-2,\dots,2,3\}$.) Если мы ещё больше расширим множество цифр, разрешив ± 3, то получим оценку $\frac{1663200}{655177} \cdot 2^n \approx 2.539 \cdot 2^n$ при правильном ответе $6 \cdot 2^n - 1$, т.е. наша оценка меньше половины от правильного числа. Если мы добавим еще одну пару цифр ± 4, получим оценку $\frac{6810804000}{2330931341} \cdot 2^n \approx 2.922 \cdot 2^n$ при правильном ответе $8 \cdot 2^n - 1$. Похоже, от расширения множества цифр оценка становится только хуже. В чём же дело?

5.2.18. (Таблица результатов для цифр $-m \dots m$.) Для того, чтобы разобраться, что происходит, сведём полученные результаты для наборов цифр $D_m = \{-m, -m+1, \dots, m\}$ с биномиальным распределением на цифрах[2] в небольшую таблицу, где $c(m)$ — это полученный из марковской цепи коэффициент c_1, участвующий в асимптотической оценке $c(m)^{-1} \cdot 2^n + O(1)$ количества n-значных чисел:

m	оценка	точное значение	$c(m)^{-1} = c_1^{-1}$	$c(m)^{-1}$	$2m$
1	$\frac{3}{2} \cdot 2^n + O(1)$	$2^{n+1} - 1$	$\frac{3}{2}$	1.5	2
2	$\frac{315}{151} \cdot 2^n + O(1)$	$2^{n+2} - 1$	$\frac{315}{151}$	2.0861	4
3	$\frac{1663200}{655177} \cdot 2^n + O(1)$	$3 \cdot 2^{n+1} - 1$	$\frac{1663200}{655177}$	2.5386	6
4	$\frac{6810804000}{2330931341} \cdot 2^n + O(1)$	$2^{n+3} - 1$	$\frac{6810804000}{2330931341}$	2.9219	8
5	$3.2608 \cdot 2^n + O(1)$	$5 \cdot 2^{n+1} - 1$	$\frac{121645100408832}{37307713155613}$	3.2608	10
16	$5.8024 \cdot 2^n + O(1)$	$2^{n+5} - 1$	5.8024	5.8024	32

(Мы добавили сюда результат при $m = 16$, однако не указали точное зна-

[1] В действительности распределение $P(2) = P(\bar{2}) = P(1) = P(\bar{1}) = 1/4$, $P(0) = 0$ даёт лучшую оценку $\frac{27}{8} \cdot 2^{-n} = 3.375 \cdot 2^n$, чем выбранное нами биномиальное распределение; однако такое распределение цифр довольно противоестественно.

[2] Судя по всему, оптимальную оценку в смысле **5.2.12** даёт не биномиальное распределение, а «экстремальное» распределение, задействующее только четыре цифры $\pm m$ и $\pm(m-1)$ с вероятностями $1/4$. Образ соответствующего распределения на D_m^ω относительно $w : D_m^\omega \to [-m, m]$ — это распределение с трапециевидной плотностью, очень близкое к равномерному при больших m. Тем не менее, мы продолжим изучать биномиальные распределения, переходящие в пределе в нормальные.

чение $c(16)^{-1}$, поскольку оно выражается дробью с восьмидесятизначными числителем и знаменателем; числитель равен $63!/78848$.)

Можно видеть, что коэффициент $c(m)^{-1}$ растёт при увеличении m гораздо медленнее, чем линейная функция. Например, при переходе от $m = 1$ к $m = 4$ он увеличивается с 1.5 до 2.922, т.е. чуть меньше, чем вдвое, в то время как «правильный» ответ увеличивается вчетверо. Это наводит на мысль, что наша оценка растёт примерно как $\sqrt{m}$, в то время как ожидаемый ответ линейно зависит от m, и расхождение только увеличивается с ростом m. И в самом деле, при переходе от $m = 1$ к $m = 2$ коэффициент $c(m)^{-1}$ увеличивается в $2.086/1.5 \approx 1.391$ раз, что очень близко к $\sqrt{2}$; при переходе от $m = 2$ к $m = 3$ — в $2.539/2.086 \approx 1.217$ раза, что близко к $\sqrt{3/2} \approx 1.225$; наконец, при переходе от $m = 3$ к $m = 4$ — в $2.922/2.539 \approx 1.151$ раза, что примерно равно $\sqrt{4/3} \approx 1.155$. Можно предположить, что при $m \to +\infty$ значение $c(m)^{-1}/\sqrt{m}$ стремится к некоторому пределу, близкому к $5.8024/\sqrt{16} \approx 1.4506$; чуть больше численных данных позволяют вычислить этот предел точнее как 1.4472. Мы увидим чуть дальше в (5.2.27.3), что на самом деле этот предел равен $\sqrt{2\pi/3}$.

5.2.19. (Правильный подсчёт количества целых чисел с n-значной записью.) Для того, чтобы объяснить получающиеся результаты, заметим, что при больших m экстремальные цифры, то есть близкие к $\pm m$, действительно стали крайне маловероятны. Это, может быть, и повысило «типичность» n-значных чисел, но зато сделало крайне маловероятным, что эти числа примут значения, близкие ко краям $-m \cdot 2^n \cdots m \cdot 2^n$, поскольку такие целые числа не могут быть записаны без крайне редких экстремальных цифр.

Это наводит на мысль о том, что теперь мы более-менее правильно вычисляем количество классов эквивалентности записей чисел длины n, но зато неправильно вычисляем другую величину, с которой мы её сравниваем. Целых чисел от $-m \cdot 2^n$ до $m \cdot 2^n$ на самом деле гораздо меньше, чем $m \cdot 2^{n+1} - 1$, потому что числа, близкие к концам этого отрезка, практически никогда не встречаются. Для того, чтобы получить логарифм их «правильного» количества, надо взять энтропию их истинного распределения, а не равномерного.

5.2.20. (Распределение целых чисел с записью в D^n, $D = \{-1, 0, 1\}$.) Например, если $D = \{-1, 0, 1\}$, т.е. $m = 1$, то целые числа ξ от $-2^n + 1$ до $2^n - 1$ получаются как значения слов из D^n не с одинаковыми веро-

ятностями, а с вероятностями

$$P(\xi = k) = \frac{2^n - |k|}{4^n} \quad \text{для целых } -2^n \le k \le 2^n \tag{5.2.20.1}$$

Здесь мы считаем, что цифры из D выбираются с естественными вероятностями $P(0) = 1/2$, $P(1) = P(\bar{1}) = 1/4$. Для того, чтобы доказать (5.2.20.1), достаточно записать цифры $\xi = \sum_{i=0}^{n-1} \xi_i 2^{n-1-i}$ в виде $\xi_i = \varepsilon_i^{(1)} - \varepsilon_i^{(2)}$, где $\varepsilon_i^{(j)}$ независимо принимают значения 0 и 1 с вероятностями $1/2$. Тогда $\xi = \xi^{(1)} - \xi^{(2)}$, где независимые $\xi^{(j)} = \sum_{i=0}^{n-1} \varepsilon_i^{(j)} 2^{n-1-i}$ принимают все целые значения от 0 до $2^n - 1$ с одинаковыми вероятностями, равными 2^{-n}.

5.2.21. (Энтропия распределения целых чисел с n-значной записью.) Теперь мы должны вычислить энтропию распределения (5.2.20.1), чтобы использовать её вместо логарифма количества чисел от -2^n+1 до 2^n-1. Эта энтропия равна

$$H(\xi) = \sum_{k=1-2^n}^{2^n-1} -\frac{2^n - |k|}{4^n} \log\Big(\frac{2^n - |k|}{4^n}\Big) \tag{5.2.21.1}$$

Заменим эту сумму на интеграл

$$\begin{aligned} H(\xi) &= \log(2^n) + 2^{-n} \sum_{k=1-2^n}^{2^n-1} -p(k/2^n) \log p(k/2^n) \\ &\approx n \log 2 + \int_{-1}^{1} -p(x) \log p(x)\, dx = n \log 2 + H(p) \end{aligned} \tag{5.2.21.2}$$

где $p(x) = (1 - |x|)[-1 \le x \le 1]$ — плотность распределения из **5.1.5**, а $H(p)$ — дифференциальная энтропия этого распределения. Таком образом, мы должны считать, что целых чисел от -2^n+1 до 2^n-1 не $2^{n+1}-1$, а приблизительно $e^{H(p)} \cdot 2^n$, где $H(p)$ — дифференциальная энтропия «треугольного» распределения вещественных чисел на $[-1, 1]$, определённого в **5.1.5**.

5.2.22. (Дифференциальная энтропия распределения $p(x) = 1 - |x|$ на $[-1, 1]$.) Теперь мы можем легко вычислить дифференциальную энтропию распределения $p(x)$ с плотностью $1 - |x|$ на $[-1, 1]$:

$$H(p) = -2 \int_0^1 x \log(x)\, dx = -x^2 \log(x)\Big|_0^1 + \int_0^1 x\, dx = \frac{1}{2} \tag{5.2.22.1}$$

Поэтому целых чисел, обладающих n-значной записью с цифрами $D = \{-1, 0, 1\}$, не $2^{n+1} - 1$, а примерно $e^{1/2} \cdot 2^n \approx 1.649 \cdot 2^n$. Отметим, что наше приближение $1.5 \cdot 2^n$, полученное с помощью цепи Маркова, не так далеко от этого значения.

5.2.23. (Дифференциальная энтропия распределения $p_4 = \chi^{*4}_{[-1/2,1/2]}$.) Что происходит при $m = 2$, $\tilde{D} = \{-2, -1, 0, 1, 2\}$? В этом случае случайная цифра $\xi_i \in \tilde{D}$ с биномиальным распределением вероятностей из **5.2.16** может быть получена как $\varepsilon_i^{(1)} + \varepsilon_i^{(2)} + \varepsilon_i^{(3)} + \varepsilon_i^{(4)}$, где $\varepsilon_i^{(j)}$ независимо и равновероятно принимают значения $\pm 1/2$. Отсюда легко находим, что плотность p_4 меры на $[-2, 2]$, являющейся образом меры-произведения $\tilde{D}^\omega$ относительно $w : \tilde{D}^\omega \to [-2, 2]$, $\xi_0\xi_1 \ldots \mapsto (0.\xi_0\xi_1 \ldots)_2 = \sum_{i=0}^{\infty} 2^{-i-1}\xi_i$, есть свёртка четырёх экземпляров $\chi_{[-1/2,1/2]}$. Поэтому $p_4(t)$ — это кусочно-кубический многочлен

$$p_4(t) = \begin{cases} 0, & \text{при } t \geq 2 \\ \frac{(2-t)^3}{6}, & \text{при } 1 \leq t \leq 2 \\ \frac{t^3}{2} - t^2 + \frac{2}{3}, & \text{при } 0 \leq t \leq 1 \\ p_4(-t), & \text{при } t < 0 \end{cases} \tag{5.2.23.1}$$

Отметим, что значения $p_4(t)$ в целых точках -2, -1, $\ldots$, 2 — это 0, $1/6$, $2/3$, $1/6$, 0; мы уже видели эту последовательность в **5.2.14**. Это совпадение неслучайно, см. (5.3.13.2).

В общем случае n-кратная свёртка $p_n = \chi^{*n}_{[-1/2,1/2]}$ задаётся формулой

$$p_n(t) = n \cdot \sum_{k=0}^{n} \frac{(-1)^k (t + \frac{n}{2} - k)_+^{n-1}}{k!(n-k)!} \tag{5.2.23.2}$$

где $x_+ = \sup(x, 0)$; эту формулу можно проверить по индукции, но проще осознать, что это выражение обладает правильной n-ой производной, либо применить к (5.2.23.2) преобразование Лапласа $\mathcal{L}$ и получить разложение $\mathcal{L}\{p_n\}(s) = \mathcal{L}\{\chi_{[-1/2,1/2]}\}(s)^n = \big((e^{s/2} - e^{-s/2})/s\big)^n$ по биному, поскольку $\mathcal{L}\{t_+^{n-1}/(n-1)!\}(s) = s^{-n}$.

Так или иначе, теперь $H(p_4)$ представляется в виде

$$H(p_4) = -2\int_0^1 \frac{x^3}{6} \log\Big(\frac{x^3}{6}\Big)\, dx - 2\int_0^1 \Big(\frac{x^3}{2} - x^2 + \frac{2}{3}\Big) \log\Big(\frac{x^3}{2} - x^2 + \frac{2}{3}\Big)\, dx \tag{5.2.23.3}$$

В принципе этот интеграл может быть выражен в замкнутой форме, поскольку интеграл $\int_a^b p(x)\log q(x)\,dx$ берётся для всех многочленов p и q; однако для этого надо находить корни многочлена $q(x)$, например, для $q(x)=x^3/2-x^2+2/3$. Итоговое выражение слишком громоздко; лучше проинтегрировать численно и получить $H(p_4)\approx 0.8667$ и $e^{H(p_4)}\approx 2.379$. Таким образом, при $m=2$ и множестве цифр $D=\{-2,-1,0,1,2\}$ мы получаем $e^{H(p_4)}\cdot 2^n \approx 2.379\cdot 2^n$ n-значных чисел, при нашей оценке с помощью цепи Маркова в $\frac{315}{151}\cdot 2^n \approx 2.0861\cdot 2^n$.

5.2.24. (Таблица новых оценок.)

m	$H(p_{2m})$	$e^{H(p_{2m})}$	$c(m)^{-1}=c_1^{-1}$	$c(m)^{-1}$	$2m$
1	$\frac{1}{2}$	1.6487	$\frac{3}{2}$	1.5	2
2	0.8667	2.3791	$\frac{315}{151}$	2.0861	4
3	1.0713	2.9192	$\frac{1663200}{655177}$	2.5386	6
4	1.2156	3.3725	$\frac{6810804000}{2330931341}$	2.9219	8
5	1.3274	3.7714	$\frac{121645100408832}{37307713155613}$	3.2608	10
16	1.9093	6.7485	5.8024	5.8024	32

Похоже, $e^{H(p_{2m})}/c(m)^{-1}$ стремится не к единице, а к какому-то другому пределу, приблизительно равному 1.165. Мы увидим чуть дальше в (5.2.28.6), что этот загадочный предел на самом деле равен $\sqrt{e/2}$.

5.2.25. (Способ избавиться от расхождения: точная интегральная формула для $c_1=c(m)$.) Отметим, что есть очень простой способ полностью избавиться от расхождения: а именно, вспомнить, что цепь Маркова даёт нам вероятность p того, что два случайных слова длины n эквивалентны, в виде $c_1 2^{-n}+O(1)$, в то время как $p=\sum_{i=1}^N p_i^2$, где p_i — вероятность того, что слово попало в i-й класс эквивалентности (см. **5.2.10**). Иначе говоря, $\{p_i\}_{1\le i\le N}$ — это в точности то самое распределение вероятностей на целых числах $-m\cdot 2^n\ldots m\cdot 2^n$, энтропию которого мы считали в **5.2.21**, а значит, следует ожидать, что $c_1=c(m)$ можно приблизительно вычислить как

$$c_1=c(m)\approx\int_{\mathbb{R}} p_{2m}(t)^2 dt \tag{5.2.25.1}$$

где теперь p_{2m} — $2m$-кратная свёртка $\chi_{[-1/2,1/2]}^{*2m}$ из (5.2.23.2). В действительности ошибка замены суммы на интеграл крайне мала и имеет порядок $O(2^{-n})$, и потому приведённая выше формула должна быть равен-

ством:

$$c(m) = c_1 = \int_{\mathbb{R}} p_{2m}(t)^2 dt \tag{5.2.25.2}$$

5.2.26. (Точная формула для $c(m)$.) Поскольку $p_{2m}(t) = p_{2m}(-t)$, равенство (5.2.25.2) можно продолжить следующим образом:

$$c(m) = \int_{\mathbb{R}} p_{2m}(t)^2 dt = (p_{2m} * p_{2m})(0) = p_{4m}(0) \tag{5.2.26.1}$$

Теперь мы можем вычислить точное значение $c(m)$ с помощью (5.2.23.2), где мы заменим индекс суммирования $k \leftrightarrow 2m - k$:

$$c(m) = p_{4m}(0) = 4m \cdot \sum_{k=1}^{2m} \frac{(-1)^k k^{4m-1}}{(2m-k)!(2m+k)!} \tag{5.2.26.2}$$

В частности, $(4m-1)! \cdot c(m) \in \mathbb{Z}$.

5.2.27. (Предел $c(m)\sqrt{m}$ при $m \to +\infty$.) Заметим, что $p_{4m}(t)$ — это плотность распределения суммы $4m$ независимых случайных величин, равнораспределённых на отрезке $[-1/2, 1/2]$, а значит, по центральной предельной теореме $p_{4m}(t)$ близко с нормальному распределению с нулевым средним и дисперсией $m/3$:

$$p_{4m}(t) \approx \frac{\sqrt{3}}{\sqrt{2\pi m}} e^{-3t^2/(2m)} \tag{5.2.27.1}$$

откуда

$$c(m) = p_{4m}(0) \approx \sqrt{\frac{3}{2\pi}} m^{-1/2} \tag{5.2.27.2}$$

так что предел $c(m)^{-1}/\sqrt{m}$ из **5.2.18** в действительности равен

$$\lim_{m \to +\infty} c(m)^{-1} m^{-1/2} = \sqrt{2\pi/3} \approx 1.4472025 \tag{5.2.27.3}$$

5.2.28. (Дифференциальная энтропия и интеграл квадрата плотности нормального распределения.) Пусть $\rho_\sigma(t)$ — плотность нормального распределения $\mathcal{N}(0, \sigma^2)$:

$$\rho_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/(2\sigma^2)} \tag{5.2.28.1}$$

Тогда

$$\int_{\mathbb{R}} \rho_\sigma(t)^2 dt = (\rho_\sigma * \rho_\sigma)(0) = \rho_{\sigma\sqrt{2}}(0) = \frac{1}{2\sigma\sqrt{\pi}} \tag{5.2.28.2}$$

и

$$\begin{aligned} H(\rho_\sigma) &= -\int_{\mathbb{R}} \rho_\sigma(t) \log \rho_\sigma(t)\, dt = \log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2}\int_{\mathbb{R}} t^2 \rho_\sigma(t)\, dt \\ &= \log(\sigma) + \log(\sqrt{2\pi e}) \end{aligned} \tag{5.2.28.3}$$

откуда

$$e^{H(\rho_\sigma)} = \sigma\sqrt{2\pi e} \tag{5.2.28.4}$$

и

$$\int_{\mathbb{R}} \rho_\sigma^2(t)\, dt = \sqrt{\frac{e}{2}} \cdot e^{-H(\rho_\sigma)} \tag{5.2.28.5}$$

Поскольку p_{2m} близко к $\mathcal{N}(0, m/6)$ при $m \gg 0$, мы можем вычислить предел

$$\lim_{m\to+\infty} e^{H(p_{2m})} c(m) = \sqrt{\frac{e}{2}} \approx 1.165822 \tag{5.2.28.6}$$

ранее возникший в **5.2.24**.

5.2.29. (Асимптотическое разложение $c(m) = p_{4m}(0)$ при $m \to +\infty$.) Точная формула (5.2.26.2) мало говорит нам о поведении $c(m) = p_{4m}(0)$ при $m \to +\infty$. Для того, чтобы изучить это поведение, вычислим $p_N(0)$ с помощью преобразования Фурье:

$$p_N(0) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{p}_N(t)\, dt = \frac{1}{2\pi}\int_{\mathbb{R}} \left(\frac{\sin(t/2)}{t/2}\right)^N dt = \frac{1}{\pi}\int_{\mathbb{R}} \operatorname{sinc}^N(t)\, dt \tag{5.2.29.1}$$

где $\operatorname{sinc}(t) = \sin(t)/t$. Запишем

$$\begin{aligned} \log\operatorname{sinc}(t) &= \log\Big(1 - \frac{t^2}{3!} + \frac{t^4}{5!} - \frac{t^6}{7!} + \frac{t^8}{9!} + O(t^{10})\Big) \\ &= -\frac{t^2}{6} - \frac{t^4}{180} - \frac{t^6}{2835} - \frac{t^8}{37800} + O(t^{10}) \end{aligned} \tag{5.2.29.2}$$

откуда

$$\operatorname{sinc}^N(t) = e^{-Nt^2/6}\Big(1 - \frac{Nt^4}{180} - \frac{Nt^6}{2835} + \frac{N^2t^8}{64800} - \frac{Nt^8}{37800} + O(t^{10})\Big) \tag{5.2.29.3}$$

Теперь мы можем ввести новую переменную $x = t\sqrt{N/3}$ и переписать интеграл (5.2.29.1) в виде

$$p_N(0) = \frac{1}{\pi\sqrt{N/3}} \int_{\mathbb{R}} e^{-x^2/2}\Big(1 - \frac{x^4}{20N} - \frac{x^6}{105N^2} + \frac{x^8}{800N^2} + O(x^8N^{-3})\Big)\, dx$$
$$= \sqrt{\frac{6}{\pi N}}\Big(1 - \frac{3}{20N} - \frac{13}{1120N^2} + O(N^{-3})\Big) \tag{5.2.29.4}$$

поскольку $\int_{\mathbb{R}} x^{2n} e^{-x^2/2} dx = (2n-1)!!\sqrt{2\pi}$. Если подставить сюда $N = 4m$, получим асимптотическое разложение $c(m)$:

$$c(m) = \sqrt{\frac{3}{2\pi}}\Big(m^{-1/2} - \frac{3}{80}m^{-3/2} - \frac{13}{17920}m^{-5/2} + O(m^{-7/2})\Big) \tag{5.2.29.5}$$

и

$$c(m)^{-1} = \sqrt{\frac{2\pi}{3}}\Big(m^{1/2} + \frac{3}{80}m^{-1/2} + \frac{191}{89600}m^{-3/2} + O(m^{-5/2})\Big) \tag{5.2.29.6}$$

Последняя формула на практике оказывается удивительно точной даже при маленьких m, как будто коэффициент при $m^{-5/2}$ в ней крайне мал, приблизительно $1/5000$.

5.3 Спектр матрицы переходов цепи Маркова

Среди загадок, рассмотренных в предыдущем пункте, осталась непроясненной как минимум одна: почему спектр матриц перехода $A = A_{m'}$ возникающих цепей Маркова, вроде (5.2.16.1) при $m' = 2$, состоит из $\{2^{-k}\}_{1\le k\le 4m'}$ и ещё одного собственного числа $2^{-4m'}$?

5.3.1. (Последствия для подсчёта вероятности возврата за n шагов.) Центральную роль в рассмотрениях предыдущего пункта играла вероятность p возврата построенной цепи Маркова в начальное состояние за n шагов. Мы получили для этой вероятности возврата формулу вида $c(m') \cdot 2^{-n} + O(4^{-n})$, и затем изучили поведение функции $c(m')$ при различных значениях m'. Однако, если бы мы знали все собственные числа λ_j матрицы переходов $A_{m'}$, то мы могли бы написать для вероятности возврата формулу вида

$$p = \sum_{j=1}^{4m'+1} c_j\lambda_j^n = \sum_{j=1}^{4m'+1} c_j(m')\lambda_j^n \tag{5.3.1.1}$$

где $\{\lambda_j\}_{1\le j\le 4m'+1}$ — собственные числа матрицы A_m в порядке убывания модуля, а $c_j = c_j(m')$ — некоторые коэффициенты, являющиеся произведениями подходящих элементов матрицы собственных векторов и обратной к ней. В частности, если верно, что $\lambda_j = 2^{-j}$ при $1 \le j \le 4m'$ и $\lambda_{4m'+1} = 2^{-4m'}$, то должна быть выполнена формула

$$p = \sum_{j=1}^{4m'} c_j 2^{-jn} + n c_{4m'+1} 2^{-4m'n} \tag{5.3.1.2}$$

для некоторых $c_j \in \mathbb{C}$. Напротив, если мы докажем существование такой формулы для всех n с *ненулевыми* коэффициентами c_j при $1 \le j \le 4m'$, отсюда будет следовать, что $\{2^{-j}\}_{1\le j\le 4m'}$ присутствуют среди собственных чисел матрицы $A_{m'}$. Оставшееся собственное число $2^{-4m'}$ тогда однозначно определяется тем, что сумма собственных чисел должна быть равна следу $\operatorname{Tr} A_{m'} = 1$.

5.3.2. (Половина коэффициентов c_j в (5.3.1.2) равны нулю.) В действительности все коэффициенты c_j в формуле (5.3.1.2) с чётными индексами j оказываются равны нулю по соображениям симметрии, как мы увидим в **5.3.16**. Например, при $m' = 2$ точная формула вероятности возврата в начальное состояние за n шагов имеет вид

$$p = \frac{151}{315}\cdot 2^{-n} + \frac{2}{9}\cdot 2^{-3n} + \frac{7}{45}\cdot 2^{-5n} + \frac{1}{7}\cdot 2^{-7n} \tag{5.3.2.1}$$

что уточняет результат **5.2.16**. Поэтому для наших целей недостаточно рассматривать только вероятности возврата в начальное состояние за n шагов. В связи с этим мы будем также изучать вероятности перехода за n шагов из любого исходного состояния s в любое конечное состояние t.

5.3.3. (Новый параметр $m = 4m'$.) Пока не поздно, введём новый параметр $m = 4m'$, поскольку исходный параметр m' (определяющий диапазон цифр $D_{m'} = \{-m', -m'+1, \dots, m'\}$ избыточной двоичной системы) почти всегда оказывается в формулах в удвоенном или в учетверённом виде. Это связано с тем, что мы выбираем очередную случайную цифру $\xi_j \in D_{m'}$ с помощью биномиального распределения, т.е. ξ_j есть сумма $2m'$ независимых случайных величин $\varepsilon_j^{(k)}$, принимающих значения $\pm 1/2$ с одинаковой вероятностью $1/2$. Цифры η_j другого случайного числа выбираются аналогичным образом, и наша марковская цепь с $4m'+1$ состояниями $s \in \{-2m', -2m'+1, \dots, 2m'\}$ переходит из состояния s в новое

состояние $t = 2s + (\xi_j - \eta_j)$, при условии, что $|t| \le 2m'$ (иначе марковская цепь перестаёт работать или переходит в специальное ошибочное состояние). Иначе говоря, мы получаем марковскую цепь, переходы $s \to t$ в которой определяются

$$t = 2s + \sum_{k=1}^{4m'} \varepsilon_j^{(k)} \tag{5.3.3.1}$$

т.е. возникают суммы $4m'$ случайных величин $\varepsilon_j^{(k)} = \pm 1/2$. Ясно, что параметр $m = 4m'$ гораздо более естественен при изучении таких марковских цепей и их матриц перехода.

5.3.4. (Марковская цепь и её матрица переходов $A = A_m$ в зависимости от нового параметра m.) Итак, теперь мы изучаем марковскую цепь с $m+1$ состоянием $-m/2, -m/2+1, \ldots, m/2$, переходы в которой задаются процессом

$$s_{j+1} = 2s_j + \sum_{k=1}^{m} \varepsilon_j^{(k)} \tag{5.3.4.1}$$

при условии, что s_{j+1} оказывается в разрешённом множестве. Заметим, что при нечётном m все состояния, начиная с s_1, будут не целыми, а полуцелыми, так что в этом случае естественно в качестве состояний марковской цепи взять полуцелые числа $-m/2, -m/2+1, \ldots, m/2$. Тем не менее, мы получаем марковскую цепь с $m+1$ состояниями вне зависимости от чётности m. Её матрица переходов $A = A_m$ есть

$$(A_m)_{st} = \begin{cases} 2^{-m}\binom{m}{t-2s+m/2}, & \text{если } |t-2s| \le m/2 \\ 0, & \text{иначе} \end{cases} \tag{5.3.4.2}$$

Здесь s и t пробегают множество состояний, т.е. целые или полуцелые числа $-m/2$, $-m/2+1$, $\ldots$, $m/2$, и $(A_m)_{st}$ — вероятность перехода из состояния s в состояние t за один шаг. Видно, что строки матрицы A — это сдвинутые и усечённые биномиальные распределения. Например, (5.2.9.1) — это теперь матрица A_4, а (5.2.16.1) — это A_8. Отметим ещё, что

$$\operatorname{Tr} A_m = \sum_{s=-m/2}^{m/2} (A_m)_{ss} = 2^{-m} \sum_{s=-m/2}^{m/2} \binom{m}{m/2-s} = 1 \tag{5.3.4.3}$$

В формулах такого вида суммы от $-m/2$ до $m/2$ будут обозначать сумму по всем целым или полуцелым индексам в указанном диапазоне в зависимости от чётности m. При желании можно преобразовать такие суммы в обычные суммы от 0 до m, если использовать $s' = s + m/2$ в качестве нового индекса суммирования.

5.3.5. (Точный подсчёт вероятности возврата в начальное состояние за n шагов.) Попробуем получить точную формулу для вероятности возврата в нулевое состояние за n шагов. Отметим, что эта задача имеет смысл только для чётного m (как в исходной постановке задачи, соответствующей $m = 4m'$), однако первые несколько этапов последующего вывода верны и для нечётного m. Согласно предыдущему пункту, состояние марковской цепи после n шагов — это случайная величина

$$\xi = \sum_{j=0}^{n-1} 2^{n-1-j} \xi_j \tag{5.3.5.1}$$

где независимые случайные величины ξ_j (изначально возникшие как разности двух случайных цифр) выбраны согласно биномиальному распределению. Иначе говоря, можно считать, что

$$\xi_j = \sum_{k=1}^{m} \varepsilon_j^{(k)} \tag{5.3.5.2}$$

где $\varepsilon_j^{(k)}$ независимо принимают значения $\pm 1/2$ с вероятностями $1/2$. В итоге мы можем записать

$$\xi = \sum_{j=0}^{n-1} 2^j \sum_{k=1}^{m} \varepsilon_j^{(k)} = \sum_{k=1}^{m} \sum_{j=0}^{n-1} 2^j \varepsilon_j^{(k)} \tag{5.3.5.3}$$

где мы заменили индекс суммирования $j \leftrightarrow n-1-j$. Заметим, что $\sum_{j=0}^{n-1} 2^j \varepsilon_j^{(k)}$ — это случайная величина $\nu_n^{(k)}$, равновероятно принимающая все 2^n полуцелых значений в диапазоне $-2^{n-1}+1/2 \ldots 2^{n-1}-1/2$ с шагом один. Таким образом, ξ — сумма m таких независимых случайных величин. Выпишем производящую функцию Ez^ξ (т.е. преобразование Фурье $Ee^{it\xi}$ с заменой $z = e^{it}$):

$$Ez^\xi = \sum_{k \in \mathbb{Z}} P(\xi = k) z^k = \left(\frac{1}{N} \sum_{j=0}^{N-1} z^{j-(N-1)/2} \right)^m \tag{5.3.5.4}$$

где мы положили $N := 2^n$. Таким образом, интересующая нас вероятность возврата $P(\xi = 0)$ — это коэффициент при z^0 в разложении Лорана рациональной функции

$$Ez^{\xi} = \left(\frac{z^{N/2} - z^{-N/2}}{N(z^{1/2} - z^{-1/2})}\right)^m = N^{-m} z^{-m(N-1)/2} \frac{(1-z^N)^m}{(1-z)^m} \tag{5.3.5.5}$$

Однако

$$(1-z)^{-m} = \sum_{\ell=0}^{+\infty} \binom{\ell+m-1}{m-1} z^{\ell} \tag{5.3.5.6}$$

и

$$(1-z^N)^m = \sum_{k=0}^{m} (-1)^k \binom{m}{k} z^{kN} \tag{5.3.5.7}$$

Отсюда мы легко находим коэффициент при z^s в Ez^{ξ}:

$$P(\xi = s) = N^{-m} \sum_{k=0}^{m} (-1)^k \binom{m}{k} \binom{s + m(N+1)/2 - 1 - kN}{m-1} \tag{5.3.5.8}$$

где сумма берётся только по тем $0 \leq k \leq m$, для которых $s + m(N+1)/2 - 1 - kN \geq m-1$, т.е. $kN \leq s + m(N-1)/2$ или $k \leq m/2 + \lfloor \frac{s-m/2}{N} \rfloor$; в последнем преобразовании мы предполагаем m чётным. В частности, при $|s| \leq m/2$ (соответствующих состояниям нашего конечного автомата и цепи Маркова) следует суммировать по $0 \leq k \leq m/2 - 1$ (если $N \geq m$), за исключением случая $s = m/2$, когда нужно учесть ещё и $k = m/2$:

$$P(\xi = s) = N^{-m} \sum_{k=[s<m/2]}^{m/2} (-1)^{m/2-k} \binom{m}{m/2-k} \binom{kN + s + m/2 - 1}{m-1} \tag{5.3.5.9}$$

где мы заменили индекс суммирования $k \leftrightarrow m/2 - k$.

Тем самым мы получили явную формулу для распределения вероятностей состояний нашей марковской цепи через n шагов, пока что при дополнительном условии, что $N = 2^n \geq m$, и для чётного m.

5.3.6. (Преобразование явной формулы вероятностей конечных состояний.) Перепишем формулу (5.3.5.9) следующим образом:

$$P(\xi = s) = \frac{m}{N^m} \sum_{k=[s<m/2]}^{m/2} (-1)^{\frac{m}{2}-k} \frac{\left(kN + s + \frac{m}{2} - 1\right)^{\underline{m-1}}}{(\frac{m}{2}-k)!(\frac{m}{2}+k)!} \tag{5.3.6.1}$$

где $x^{\underline{n}} = x(x-1)\cdots(x-n+1)$ — n-ая убывающая степень x, и мы по-прежнему предполагаем m чётным. Поскольку

$$x^{\underline{n}} = \sum_{\ell=0}^{n} (-1)^{n-\ell} \begin{bmatrix} n \\ \ell \end{bmatrix} x^{\ell} \tag{5.3.6.2}$$

где $\begin{bmatrix} n \\ \ell \end{bmatrix}$ обозначает число Стирлинга первого рода (количество подстановок n элементов, раскладывающихся в ℓ циклов), мы можем переписать сумму в (5.3.6.1) как многочлен от $N = 2^n$ (при чётном m):

$$P(\xi = s) = \frac{m}{N^m} \sum_{k=[s<m/2]}^{m/2} \sum_{\ell=0}^{m-1} \frac{(-1)^{\frac{m}{2}-k+\ell+1}}{(\frac{m}{2}-k)!(\frac{m}{2}+k)!} \begin{bmatrix} m-1 \\ \ell \end{bmatrix} \Big(kN + s + \frac{m}{2} - 1\Big)^{\ell} \tag{5.3.6.3}$$

Видно, что эта сумма может быть записана как некоторый многочлен $Q_m(s; N)$ от N степени $m-1$, который в действительности является многочленом от s и N, если ограничиться только значениями $-m/2 \le s < m/2$. Поэтому $P(\xi = s)$ представима в виде

$$P(\xi = s) = \sum_{j=1}^{m} c_{s,j} N^{-j} = \sum_{j=1}^{m} c_{s,j} 2^{-jn} \tag{5.3.6.4}$$

с рациональными $c_{s,j}$ — коэффициентами многочлена $Q_m(s; N)$. Это формула вида (5.3.1.2), её существование как раз и означает, что $\{2^{-j}\}_{1\le j\le m}$ — это собственные числа матрицы A_m. Однако нам надо как-то убедиться, что для каждого j есть хотя бы один ненулевой $c_{s,j}$.

5.3.7. (Явная формула для всех элементов A_m^n.) Заметим, что мы можем аналогичным образом вычислить все вероятности перехода $t \to s$ за n шагов, т.е. все элементы $(A_m^n)_{ts}$, $-m/2 \le s, t \le m/2$, причём вне зависимости от чётности m. Рассуждая как при выводе (5.3.6.1), получаем

$$(A_m^n)_{ts} = P(\xi = s - Nt) = \frac{m}{N^m} \sum_{k=[s<m/2]}^{\frac{m}{2}-t} (-1)^{\frac{m}{2}-t-k} \frac{\big(kN + s + \frac{m}{2} - 1\big)^{\underline{m-1}}}{(\frac{m}{2}-t-k)!(\frac{m}{2}+t+k)!} \tag{5.3.7.1}$$

при условии $N = 2^n \ge m$. В этой формуле важно, что $s + m/2$ и $t + m/2$ — всегда целые числа в диапазоне от 0 до m. Кроме того, при нечётном m

существенно, что $N = 2^n$ чётно. Мы можем снова разложить $x^{\underline{m-1}}$ при помощи чисел Стирлинга:

$$(A_m^n)_{ts} = \frac{m}{N^m} \sum_{k=[s<m/2]}^{m/2-t} \sum_{\ell=0}^{m-1} \frac{(-1)^{\frac{m}{2}-t-k+m-\ell-1} \left[{m-1 \atop \ell}\right]}{(\frac{m}{2}-k-t)!(\frac{m}{2}+k+t)!} \Big(kN + \Big(\frac{m}{2}+s-1\Big)\Big)^{\ell} \tag{5.3.7.2}$$

Теперь разложим $(kN + (m/2 + s - 1))^{\ell}$ по биному:

$$(A_m^n)_{ts} = \frac{m}{N^m} \sum_{k=[s<m/2]}^{m/2-t} \sum_{\ell=0}^{m-1} \sum_{j=0}^{\ell} \frac{(-1)^{\frac{m}{2}-t-k+m-\ell-1} \left[{m-1 \atop \ell}\right] \binom{\ell}{j}}{(\frac{m}{2}-t-k)!(\frac{m}{2}+t+k)!} k^j N^j \Big(\frac{m}{2}+s-1\Big)^{\ell-j} \tag{5.3.7.3}$$

5.3.8. (Разложение матрицы A_m^n.) В итоге мы можем записать

$$A_m^n = \sum_{j=0}^{m-1} N^{-j-1} \boldsymbol{C}_j^{(m)} \tag{5.3.8.1}$$

где по-прежнему $N = 2^n$, и

$$\begin{aligned} (\boldsymbol{C}_{m-1-j}^{(m)})_{ts} = m \sum_{k=[s<m/2]}^{m/2-t} & \frac{(-1)^{\frac{m}{2}-t-k} k^j}{(\frac{m}{2}-t-k)!(\frac{m}{2}+t+k)!} \\ & \times \sum_{\ell=j}^{m-1} (-1)^{m-\ell-1} \left[{m-1 \atop \ell}\right] \binom{\ell}{j} \Big(\frac{m}{2}+s-1\Big)^{\ell-j} \end{aligned} \tag{5.3.8.2}$$

при $0 \le j \le m-1$. Удивительным (на самом деле — ожидаемым) образом вторая сумма не зависит от t, а первая зависит от s только нижним индексом суммирования $[s < m/2]$, который добавляет дополнительное слагаемое с $k = 0$ при $s = m/2$. Однако это слагаемое содержит множитель $k^j = 0^j$ и потому существенно только при $j = 0$. Таким образом, при $j < m-1$ матрица $\boldsymbol{C}_j^{(m)}$ оказывается матрицей ранга один, а $\boldsymbol{C}_{m-1}^{(m)}$ — матрица ранга два.

5.3.9. (Разложения матриц $\boldsymbol{C}_j^{(m)}$.) Мы видим, что матрицы $\boldsymbol{C}_j^{(m)}$ при $0 \le j \le m-2$ являются матрицами ранга один и потому могут быть записаны как произведение вектора на ковектор, т.е. столбца на строку: $\boldsymbol{C}_j^{(m)} = \boldsymbol{v}_j \boldsymbol{u}_j^T$. Более того, формула (5.3.8.2) подсказывает, как именно

можно выбрать эти вектор и ковектор:

$$(\boldsymbol{v}_{m-1-j})_t = m\binom{m-1}{j}\sum_{k=1}^{m/2-t}\frac{(-1)^{m/2-t-k}k^j}{(\frac{m}{2}-t-k)!(\frac{m}{2}+t+k)!} \tag{5.3.9.1}$$

$$(\boldsymbol{u}_{m-1-j})_s = \binom{m-1}{j}^{-1}\sum_{\ell=j}^{m-1}(-1)^{m-\ell-1}\begin{bmatrix}m-1\\ \ell\end{bmatrix}\binom{\ell}{j}\Big(\frac{m}{2}+s-1\Big)^{\ell-j} \tag{5.3.9.2}$$

где по-прежнему $-m/2 \le s,t \le m/2$ и $0 \le j \le m-1$, и мы злоупотребили возможностью домножения $\boldsymbol{v}_{m-1-j}$ на произвольно выбранные нормировочные множители $\binom{m-1}{j}$ с одновременным делением $\boldsymbol{u}_{m-1-j}$ на те же множители. При $j=0$ мы обозначим получающийся вектор $\boldsymbol{v}_{m-1}$ также через $\boldsymbol{v}^+_{m-1}$, чтобы подчеркнуть, что мы начинаем суммирование с $k=1$, а не с $k=0$. Тогда вариант, соответствующий началу суммирования с $k=0$, можно обозначить $\boldsymbol{v}^-_{m-1}$. Кроме того, при $j=0$ нам понадобятся дополнительные вектор и ковектор, соответствующие дополнительному слагаемому с $k=0$ в (5.3.8.2), которое надо прибавить при $s=m/2$:

$$(\tilde{\boldsymbol{v}}_m)_t = (\boldsymbol{v}^-_{m-1})_t - (\boldsymbol{v}^+_{m-1})_t = \frac{(-1)^{m/2-t}\cdot m}{(\frac{m}{2}-t)!(\frac{m}{2}+t)!} = \frac{(-1)^{m/2-t}}{(m-1)!}\binom{m}{m/2-t} \tag{5.3.9.3}$$

$$(\tilde{\boldsymbol{u}}_m)_s = (\boldsymbol{u}_{m-1})_s\cdot[s\ge m/2] = (\boldsymbol{u}_{m-1})_{m/2}\cdot\delta_{s,m/2} = (m-1)!\delta_{s,m/2} \tag{5.3.9.4}$$

где мы сразу подставили значение $(\boldsymbol{u}_{m-1})_{m/2} = (m-1)!$, которое будет вычислено в (5.3.20.6). Теперь

$$\boldsymbol{C}^{(m)}_j = \begin{cases}\boldsymbol{v}_j\boldsymbol{u}_j^T, & \text{при } 0\le j\le m-2\\ \boldsymbol{v}^+_{m-1}\boldsymbol{u}^T_{m-1}+\tilde{\boldsymbol{v}}_m\tilde{\boldsymbol{u}}^T_m, & \text{при } j=m-1\end{cases} \tag{5.3.9.5}$$

и

$$A^n_m = \sum_{j=0}^{m-1}2^{-(j+1)n}\boldsymbol{C}^{(m)}_j = \sum_{j=0}^{m-1}2^{-(j+1)n}\boldsymbol{v}_j\boldsymbol{u}_j^T + 2^{-mn}\tilde{\boldsymbol{v}}_m\tilde{\boldsymbol{u}}^T_m \tag{5.3.9.6}$$

Строго говоря, мы пока что доказали последнее равенство только при $N=2^n\ge m$, но мы вскоре увидим в **5.3.11**, что оно верно для всех целых n.

Лемма 5.3.10 *(Проекторы на собственные подпространства.) Пусть* $\boldsymbol{A}$, $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_m$ *— квадратные матрицы одинакового размера* $N \times N$ *над полем* k*, и пусть* $\lambda_1, \ldots, \lambda_m \in k$ *— различные скаляры. Предположим, что*

$$\boldsymbol{A}^n = \sum_{j=1}^{m} \lambda_j^n \boldsymbol{C}_j \quad \text{для всех целых } n \geq n_0 \tag{5.3.10.1}$$

для некоторого $n_0 \in \mathbb{N}_0$*. Тогда:*

a) Минимальный многочлен матрицы $\boldsymbol{A}$ *делит* $p(X) := X^{n_0} \prod_{j=1}^{m} (X - \lambda_j)$.

b) $\operatorname{spec} \boldsymbol{A} \subset \{0, \lambda_1, \ldots, \lambda_m\}$*. Если* $n_0 = 0$*, то* $\operatorname{spec} \boldsymbol{A} \subset \{\lambda_1, \ldots, \lambda_m\}$.

c) Собственные подпространства $V_\lambda \subset V = k^N$ *для всех ненулевых собственных чисел* λ *матрицы* $\boldsymbol{A}$ *совпадают с корневыми подпространствами. Если* $n_0 \leq 1$*, это верно и для нулевого собственного числа.*

d) Пусть $n_0 \geq 1$ *и все* $\lambda_i \neq 0$*. Положим* $\boldsymbol{C}_0 := 1 - \sum_{j=1}^{m} \boldsymbol{C}_j$ *и* $\lambda_0 := 0$*. Тогда* $\{\boldsymbol{C}_j\}_{0 \leq j \leq m}$ *— коммутирующие ортогональные идемпотенты, коммутирующие также и с матрицей* $\boldsymbol{A}$ *и выражающиеся как многочлены от* $\boldsymbol{A}$*. Это семейство ортогональных идемпотентов определяет разложение* $V = k^N$ *в прямую сумму корневых подпространств, соответствующих собственным числам* λ_j *матрицы* $\boldsymbol{A}$.

e) Если (5.3.10.1) *выполнено при* $n \geq n_0$*, все* $\lambda_i \neq 0$*, и матрица* $\boldsymbol{A}$ *обратима, то* (5.3.10.1) *выполнено для всех* $n \geq 0$*, и даже для всех* $n \in \mathbb{Z}$.

Доказательство. a) Заметим, что для любого многочлена $P(X)$, делящегося на X^{n_0}, верно равенство $P(\boldsymbol{A}) = \sum_{j=1}^{m} P(\lambda_j) \boldsymbol{C}_j$, являющееся линейной комбинацией равенств вида (5.3.10.1). Взяв в качестве $P(X)$ многочлен $X^{n_0}(X - \lambda_1) \cdots (X - \lambda_m)$, получаем $P(\boldsymbol{A}) = 0$, т.е. $P(X)$ делится на минимальный многочлен матрицы $\boldsymbol{A}$.

b) Собственные числа матрицы $\boldsymbol{A}$ — это корни её минимального многочлена, так что $\operatorname{spec} \boldsymbol{A}$ содержится во множестве корней $p(X)$ из a).

c) В самом деле, из a) следует, что все ненулевые корни минимального многочлена матрицы $\boldsymbol{A}$ — простые. Это как раз означает, что соответствующие корневые подпространства совпадают с собственными. При $n_0 \leq 1$ это же верно и для нулевого собственного числа.

d) Пусть $Q_k(X) = c_k X^{n_0} \prod_{j\neq k}(X-\lambda_j)$ — многочлен, делящийся на X^{n_0}, и такой, что $Q_k(\lambda_j) = \delta_{jk}$. Поскольку $Q_k(X)$ делится на X^{n_0}, снова $Q_k(\boldsymbol{A}) = \sum_{j=1}^m Q_k(\lambda_j)\boldsymbol{C}_j = \boldsymbol{C}_k$. Таким образом, $\boldsymbol{C}_j$ при $1 \leq j \leq m$ являются многочленами от $\boldsymbol{A}$, а значит, это верно и для $\boldsymbol{C}_0 := \boldsymbol{E}_N - \sum_{j=1}^m \boldsymbol{C}_j$. Далее, $\boldsymbol{C}_j\boldsymbol{C}_k = Q_j(\boldsymbol{A})Q_k(\boldsymbol{A}) = (Q_jQ_k)(\boldsymbol{A}) = \delta_{jk}\boldsymbol{C}_j$, поскольку Q_jQ_k делится на X^{n_0}. Это доказывает, что $\{\boldsymbol{C}_j\}_{1\leq j\leq m}$ — коммутирующие ортогональные идемпотенты, а значит, если их дополнить $\boldsymbol{C}_0 = \boldsymbol{E}_N - \sum_{j=1}^m \boldsymbol{C}_j$, мы получим полное семейство ортогональных идемпотентов, коммутирующих с матрицей $\boldsymbol{A}$. Это семейство определяет разложение $V = k^N$ в прямую сумму $V = \bigoplus_{0\leq j\leq m} V_j$, где $V_j := \boldsymbol{C}_j(V)$, причём это разложение согласовано с $\boldsymbol{A}$. Теперь легко видеть, что ограничение $\boldsymbol{A}$ на V_j совпадает с умножением на скаляр λ_j — например, потому что $(\boldsymbol{A} - \lambda_j)\boldsymbol{C}_j = (\boldsymbol{A}-\lambda_j)Q_j(\boldsymbol{A}) = 0$, так как $(X-\lambda_j)Q_j(X)$ делится на минимальный многочлен матрицы $\boldsymbol{A}$. Аналогично, $\boldsymbol{A}^{n_0}\boldsymbol{C}_0 = \boldsymbol{A}^{n_0}Q_0(\boldsymbol{A}) = 0$, где $Q_0(X) := 1 - \sum_{j=1}^m Q_j(X)$, потому что $X^{n_0}Q_0(X)$ делится на минимальный многочлен матрицы $\boldsymbol{A}$. Это доказывает, что V_j — действительно корневые подпространства матрицы $\boldsymbol{A}$, соответствующие различным собственным числам λ_j.

e) Это следует из того, что $\boldsymbol{C}_j$ являются проекторами на собственные пространства матрицы $\boldsymbol{A}$, отвечающие λ_j, причём $\boldsymbol{C}_0$ из пункта d) обязательно равно нулю, так как нуль по предположению не является собственным числом матрицы $\boldsymbol{A}$.

5.3.11. (Завершение вычисления спектра матрицы переходов $A = A_m$.) Заметим, что (5.3.9.6) вместе с **5.3.10**,b) доказывают, что собственные числа матрицы $A = A_m$ содержатся во множестве 0, 2^{-1}, 2^{-2}, …, 2^{-m}. При этом ранги матриц $\boldsymbol{C}_j^{(m)}$ из (5.3.9.6) не превосходят единицы при $0 \leq j < m-1$ и двойки при $j = m-1$. Поэтому кратности собственных чисел 2^{-j} при $1 \leq j \leq m$ не превышают единицы (но в принципе могут быть и нулями), а 2^{-m} может иметь кратность два. С другой стороны, собственных чисел с учётом кратности должно быть $m+1$ штук, и их сумма равна $\operatorname{Tr} A_m = 1$ согласно (5.3.4.3). Единственный способ, как такое может быть выполнено — это если все эти собственные числа присутствуют в спектре матрицы A_m с максимальной разрешённой кратностью, т.е. 2^{-j} при $1 \leq j < m$ с кратностью один и 2^{-m} с кратностью два. Поскольку это уже даёт $m+1$ собственное число, ноль не является собственным числом матрицы A_m. Мы доказали, что *собственные числа матрицы переходов A_m — это в точности $\{2^{-j}\}_{1\leq j\leq m-1}$ с*

кратностью один и 2^{-m} *с кратностью два.* В частности, *матрица* A_m *невырождена.* Кроме того, поскольку матрицы $\boldsymbol{C}_j^{(m)}$ являются проекторами на соответствующие собственные пространства, их ранги должны быть равны кратностям соответствующих собственных чисел, т.е. *ранг матрицы* $\boldsymbol{C}_j^{(m)}$ *равен в точности единице при* $0 \leq j \leq m-2$ *и двойке при* $j = m-1$. Отсюда и из разложений матриц $\boldsymbol{C}_j^{(m)}$ из (5.3.9.5) следует, что *вектора* $\boldsymbol{v}_j$ *и* $\tilde{\boldsymbol{v}}_m$ *образуют собственный базис матрицы* A*, а* $\boldsymbol{u}_j$ *и* $\tilde{\boldsymbol{u}}_m$ *— транспонированной матрицы* A^T (см. **5.3.12** ниже).

Кроме того, **5.3.10**,e) доказывает, что *равенство* (5.3.9.6) *выполнено при всех целых* n.

5.3.12. ($\boldsymbol{v}_j$ и $\boldsymbol{u}_j$ — собственные вектора и ковектора A.) Заметим, что

$$\sum_{j=0}^{m-1} \boldsymbol{C}_j^{(m)} = \boldsymbol{E}_{m+1} \tag{5.3.12.1}$$

поскольку мы только что видели, что (5.3.9.6) выполнено и для $n = 0$. Кроме того, разложения (5.3.9.5) позволяют переписать это равенство в виде

$$\boldsymbol{v}_0\boldsymbol{u}_0^T + \cdots + \boldsymbol{v}_{m-1}^+\boldsymbol{u}_{m-1}^T + \tilde{\boldsymbol{v}}_m\tilde{\boldsymbol{u}}_m^T = \boldsymbol{E}_{m+1} \tag{5.3.12.2}$$

т.е. мы разложили единичную матрицу размера $m+1$ в сумму ровно $m+1$ разложимого тензора. Такое может быть, только если все $\boldsymbol{v}_j$ (вместе со временно переобозначенным $\boldsymbol{v}_m := \tilde{\boldsymbol{v}}_m$; напомним, что $\boldsymbol{v}_{m-1} = \boldsymbol{v}_{m-1}^+$) образуют базис векторного пространства $V = \mathbb{Q}^{m+1}$, а $\boldsymbol{u}_j$ — двойственный базис. В частности, должны быть выполнены соотношения ортогональности

$$\boldsymbol{u}_i^T\boldsymbol{v}_j = \delta_{ij} \quad \text{для } 0 \leq i, j \leq m \tag{5.3.12.3}$$

где мы снова считаем, что $\boldsymbol{u}_m = \tilde{\boldsymbol{u}}_m$ и $\boldsymbol{v}_m = \tilde{\boldsymbol{v}}_m$. Более того, теперь разложения (5.3.9.5) означают, что $\boldsymbol{v}_j$ — базис образа проектора $\boldsymbol{C}_j^{(m)}$, т.е. собственного пространства, соответствующего 2^{-j-1}, при $0 \leq j \leq m-2$, а $\boldsymbol{v}_{m-1}^+$ и $\tilde{\boldsymbol{v}}_m$ — базис собственного пространства, соответствующего 2^{-m}. Иначе говоря, $\boldsymbol{v}_j$ — это собственный базис матрицы A_m, а $\boldsymbol{u}_j$ — двойственный к нему собственный базис транспонированной матрицы A_m^T. Если V — матрица со столбцами $\boldsymbol{v}_j$, $0 \leq j \leq m$, а U — матрица со столбцами $\boldsymbol{u}_j$, то (5.3.12.2) означает, что

$$VU^T = \boldsymbol{E}_{m+1} \tag{5.3.12.4}$$

т.е. $U = V^{-T}$ (откуда снова следует, что $U^T V = \boldsymbol{E}_{m+1}$, т.е. (5.3.12.3)), и (5.3.9.6) может быть записано как

$$A_m^n = V D^n U^T = V D^n V^{-1} \quad , \quad \text{где } D = \operatorname{diag}(2^{-1}, 2^{-2}, \ldots, 2^{-m}, 2^{-m}) \tag{5.3.12.5}$$

5.3.13. (Интерпретация координат собственных векторов $\boldsymbol{v}_j$.) Поскольку вектора $\boldsymbol{v}_j$ из (5.3.9.1) оказались собственными векторами матрицы A_m, полезно разобраться с тем, каковы их координаты, как их проще всего вычислить, и каково их асимптотическое поведение при $m \to +\infty$. Начнём с того, что формула (5.3.9.1) очень похожа на формулу (5.2.23.2), задающую кусочно-полиномиальную плотность $p_m(t)$ m-кратной свёртки $\chi_{[-1/2,1/2]}^{*m}$ равномерного распределения на $[-1/2, 1/2]$:

$$p_m(t) = m \cdot \sum_{k=0}^{m} \frac{(-1)^k (t + m/2 - k)_+^{m-1}}{k!(m-k)!} \tag{5.3.13.1}$$

Отсюда мы сразу (после замены $t \leftrightarrow -t$ и затем $k \leftrightarrow m/2 - t - k$) получаем, что

$$(\boldsymbol{v}_0)_t = p_m(-t) = p_m(t) \quad \text{при } t = -\frac{m}{2}, -\frac{m}{2}+1, \ldots, \frac{m}{2} \tag{5.3.13.2}$$

поскольку $p_m(t)$ — четная функция. Более общо, j-кратное дифференцирование формулы (5.3.13.1) даёт

$$(\boldsymbol{v}_j)_t = \frac{(-1)^j}{j!} p_m^{(j)}(t) \tag{5.3.13.3}$$

при $0 \le j \le m-2$ и $-m/2 \le t \le m/2$; здесь нам пригодился нормировочный множитель $\binom{m-1}{j}$ из (5.3.9.1), который удачно сократился. При $j = m-1$ ситуация чуть более сложная, так как $(m-1)$-ая производная функции $(t + m/2 - k)_+^{m-1}$ разрывна в (полу)целых точках, и нам надо выбрать, хотим ли мы использовать левостороннюю или правостороннюю производную, что соответствует началу суммирования с $k = 0$ или $k = 1$ в формуле (5.3.7.1). В итоге получаем

$$(\boldsymbol{v}_{m-1}^+)_t = \frac{(-1)^{m-1}}{(m-1)!} p_m^{(m-1)}(t+) \tag{5.3.13.4}$$

(для проверки заметим, что $(\boldsymbol{v}_{m-1}^+)_t = 0$ при $t = m/2$ согласно (5.3.9.1), так что правильный выбор именно таков), т.е. формула (5.3.13.3) верна

и при $j = m-1$, если условиться использовать правосторонние производные. Аналогично, $\boldsymbol{v}^-_{m-1}$ возникает при использовании левосторонних производных:

$$(\boldsymbol{v}^-_{m-1})_t = \frac{(-1)^{m-1}}{(m-1)!} p_m^{(m-1)}(t-) \tag{5.3.13.5}$$

При этом разность между двумя вариантами $(m-1)$-ой производной даёт вспомогательный вектор $\tilde{\boldsymbol{v}}_m = \boldsymbol{v}^-_{m-1} - \boldsymbol{v}^+_{m-1}$ из (5.3.9.3):

$$(\tilde{\boldsymbol{v}}_m)_t = \frac{(-1)^{m/2-t}}{(m-1)!}\binom{m}{m/2-t} = \frac{(-1)^m}{(m-1)!}\big(p_m^{(m-1)}(t+) - p_m^{(m-1)}(t-)\big) \tag{5.3.13.6}$$

Таким образом, *с точностью до умножения на константу собственный вектор $\boldsymbol{v}_{j+1}$ — это таблица значений j-ой производной кусочно-полиномиальной функции $p_m(t)$ в целых или полуцелых точках $t = -m/2$, $-m/2+1$, …, $m/2$. При $j = m-1$ надлежит использовать правосторонние производные, а вспомогательный вектор $\tilde{\boldsymbol{v}}_m$ есть таблица величин разрывов $(m-1)$-ой производной функции $p_m(t)$.*

5.3.14. (Симметричность собственных векторов $\boldsymbol{v}_j$.) Поскольку $p_m(t)$ — чётная функция, мы сразу получаем из (5.3.13.3), что

$$(\boldsymbol{v}_j)_{-t} = (-1)^j (\boldsymbol{v}_j)_t \quad \text{при } |t| \le m/2,\ 0 \le j \le m-2 \tag{5.3.14.1}$$

что неочевидно из формулы (5.3.9.1), хотя и вполне ожидаемо, поскольку матрица A_m коммутирует с «оператором чётности» $P : \boldsymbol{e}_t \mapsto \boldsymbol{e}_{-t}$, $P = (\delta_{s,-t})$, а значит, собственные вектора $\boldsymbol{v}_j$ матрицы A_m, соответствующие простым собственным числам, должны быть и собственными векторами P, т.е. равенство $P\boldsymbol{v}_j = \pm\boldsymbol{v}_j$ при $0 \le j \le m-2$ было ожидаемо. Что же касается $j = m-1$, там ситуация иная из-за разницы между левосторонними и правосторонними производными, см. (5.3.13.4) и (5.3.13.5). Мы получаем

$$(-1)^{m-1} P\boldsymbol{v}^+_{m-1} = \boldsymbol{v}^-_{m-1} = \boldsymbol{v}^+_{m-1} + \tilde{\boldsymbol{v}}_m \tag{5.3.14.2}$$

$$(-1)^{m-1} P(\boldsymbol{v}^+_{m-1} + \tilde{\boldsymbol{v}}_m) = (-1)^{m-1} P\boldsymbol{v}^-_{m-1} = \boldsymbol{v}^+_{m-1} \tag{5.3.14.3}$$

откуда можно заключить, что

$$P(\boldsymbol{v}^+_{m-1} + \tilde{\boldsymbol{v}}_m/2) = (-1)^{m-1}(\boldsymbol{v}^+_m + \tilde{\boldsymbol{v}}_m/2) \tag{5.3.14.4}$$

$$P\tilde{\boldsymbol{v}}_m = (-1)^m \tilde{\boldsymbol{v}}_m \tag{5.3.14.5}$$

Иначе говоря, если мы хотим получить совместный собственный базис коммутирующих матриц A_m и P, нам надлежит в качестве собственных векторов, соответствующих собственному числу 2^{-m} матрицы A_m, использовать

$$\boldsymbol{v}'_{m-1} = \frac{\boldsymbol{v}^+_{m-1} + \boldsymbol{v}^-_{m-1}}{2} = \boldsymbol{v}^+_{m-1} + \frac{\tilde{\boldsymbol{v}}_m}{2} \tag{5.3.14.6}$$

$$\boldsymbol{v}'_m = \boldsymbol{v}^-_{m-1} - \boldsymbol{v}^+_{m-1} = \tilde{\boldsymbol{v}}_m \tag{5.3.14.7}$$

а также

$$\boldsymbol{u}'_{m-1} = \boldsymbol{u}_{m-1} \tag{5.3.14.8}$$

$$\boldsymbol{u}'_m = \tilde{\boldsymbol{u}}_m - \frac{\boldsymbol{u}_{m-1}}{2} \tag{5.3.14.9}$$

чтобы сохранить разложение $\boldsymbol{C}^{(m)}_{m-1} = \boldsymbol{v}'_{m-1}\boldsymbol{u}'^T_{m-1} + \boldsymbol{v}'_m\boldsymbol{u}'^T_m$ из (5.3.9.5), обеспечивающее двойственность базисов.

Тогда $P\boldsymbol{v}'_{m-1} = (-1)^{m-1}\boldsymbol{v}'_{m-1}$ и $P\boldsymbol{v}'_m = (-1)^m\boldsymbol{v}'_m$, откуда следует $P\boldsymbol{u}'_{m-1} = (-1)^{m-1}\boldsymbol{u}'_{m-1}$ и $P\boldsymbol{u}'_m = (-1)^m\boldsymbol{u}'_m$ для двойственного базиса. Вектор $\boldsymbol{v}'_{m-1}$ соответствует выбору полусуммы левосторонней и правосторонней производной в точках разрыва, примерно как это обычно делается в теории рядов Фурье. У нас возникает соблазн переобозначить «основной» собственный вектор $\boldsymbol{v}_{m-1}$, выбрав его равным $\boldsymbol{v}'_{m-1}$, а не $\boldsymbol{v}^+_{m-1}$, и оставить вспомогательный $\tilde{\boldsymbol{v}}_m$, возможно, переобозначив его $\boldsymbol{v}_m$. Тогда $\{\boldsymbol{v}_j\}_{0\le j\le m}$ будет совместным собственным базисом A_m и P, причём $P\boldsymbol{v}_j = (-1)^j\boldsymbol{v}_j$ для всех j.

5.3.15. (Интерпретация $\boldsymbol{v}'_{m-1}$ и $\boldsymbol{v}'_m$ как усреднённых производных.) Отметим, что $\boldsymbol{v}'_{m-1}$ и $\boldsymbol{v}'_m = \tilde{\boldsymbol{v}}_m = \boldsymbol{v}_m$ также допускают интерпретацию как производные $p_m(t)$, усреднённые по маленьким отрезкам с центром в t:

$$(\boldsymbol{v}'_{m-1})_t = \frac{2}{\varepsilon}\lim_{\varepsilon\to 0}\int_{t-\varepsilon}^{t+\varepsilon}\frac{p_m^{(m-1)}(t)}{(m-1)!}\,dt \tag{5.3.15.1}$$

$$(\boldsymbol{v}'_m)_t = \lim_{\varepsilon\to 0}\int_{t-\varepsilon}^{t+\varepsilon}\frac{p_m^{(m)}(t)}{(m-1)!}\,dt \tag{5.3.15.2}$$

где последний интеграл надлежит понимать как интеграл обобщённых функций. Кроме того, у этих двух векторов «правильная» чётность, соответствующая $(m-1)$-ой и m-ой производным чётной функции $p_m(t)$.

5.3.16. (Последствия симметричности $\boldsymbol{v}_j$ и $\boldsymbol{u}_j$ для коэффициентов c_j формулы (5.3.1.2).) Заметим, что коэффициенты c_j формулы (5.3.1.2) для точного вычисления вероятности возврата в начальное состояние за n шагов — это числа

$$c_j = (\boldsymbol{v}_{j-1})_0 \cdot (\boldsymbol{u}_{j-1})_0 \tag{5.3.16.1}$$

Поскольку $P\boldsymbol{v}_{j-1} = -\boldsymbol{v}_{j-1}$ для чётных j согласно **5.3.14**, мы получаем $(\boldsymbol{v}_{j-1})_0 = 0$ и как следствие $c_j = 0$ для всех чётных j в (5.3.1.2).

5.3.17. (Эффективное вычисление всех собственных векторов.) Поскольку производные многочлена — это (с точностью до умножения на факториал) его коэффициенты после замены переменной, можно эффективно вычислить все собственные вектора $\boldsymbol{v}_j$ с помощью следующего алгоритма, строящего последовательность многочленов $P_t(X) \in \mathbb{Z}[X]$, таких, что $P_t(x) = (m-1)!\cdot p_m(t+x)$ при $0 < x < 1$; коэффициенты этих многочленов сразу дают $(\boldsymbol{v}_j)_t$:

S0) $P_{-m/2-1}(X) \leftarrow 0$, $t \leftarrow -m/2$

S1) $c_t \leftarrow (-1)^{m/2+t}\binom{m}{m/2+t}$, $(\tilde{\boldsymbol{v}}_m)_t = (-1)^m c_t/(m-1)!$

S2) $P_t(X) \leftarrow P_{t-1}(X+1) + c_t X^{m-1}$

S3) $(\boldsymbol{v}_j)_t = \frac{(-1)^j}{(m-1)!}[X^j]P_t(X)$ для $j = 0 \ldots m-1$. Затем $t \leftarrow t+1$, и если $t \le m/2$, вернуться к шагу S1).

Здесь $[X^j]P_t(X)$ обозначает коэффициент при X^j в $P_t(X)$. Самый трудоёмкий шаг в этом алгоритме — это S2, который требует $O(m^2)$ арифметических операций (целочисленных сложений) для замены переменной в многочлене. Если мы хотим вычислить симметризованный вектор $\boldsymbol{v}'_{m-1}$ вместо $\boldsymbol{v}^+_{m-1}$, надо слегка исправить S3), вычитая $c_t/2$ из $[X^j]P_t(X)$ при $j = m-1$.

5.3.18. (Связь собственных векторов $\boldsymbol{v}_j$ с многочленами Эрмита.) Заметим, что $p_m(t)$ — это плотность m-кратной свёртки $\chi^{*m}_{[-1/2,1/2]}$, и потому по центральной предельной теореме $p_m(t)$ приблизительно равно плотности $\rho_\sigma(t)$ нормального распределения $\mathcal{N}(0, m/12)$, где $\sigma = \sqrt{m/12}$:

$$p_m(t) \approx \rho_\sigma(t) = \sqrt{\frac{6}{\pi m}}\, e^{-6t^2/m} \tag{5.3.18.1}$$

Поскольку $(\boldsymbol{v}_j)_t$ задаются значениями j-ой производной $(-1)^j p_m^{(j)}(t)/j!$, мы можем предположить, что

$$(\boldsymbol{v}_j)_t \approx \frac{(-1)^j}{j!}\rho_\sigma^{(j)}(t) \tag{5.3.18.2}$$

Однако j-ая производная $e^{-t^2/2}$ равна $(-1)^j h_j(t)e^{-t^2/2}$, где $h_j(t)$ — многочлен Эрмита степени j. Поэтому вектор $\boldsymbol{v}_j$ (по крайней мере, при маленьких j) имеет отношение к многочлену Эрмита $h_j(t)$, или скорее к его отмасштабированной версии $h_j^{[m/12]}(t)$:

$$(\boldsymbol{v}_j)_t \approx \frac{1}{j!\sqrt{2\pi}}\Big(\frac{m}{12}\Big)^{-j-1/2} h_j^{[m/12]}(t)e^{-6t^2/m} \tag{5.3.18.3}$$

где

$$h_n^{[\sigma^2]}(X) = \sigma^n h_n(\sigma^{-1}X) = \sum_{r=0}^{\lfloor n/2\rfloor} \frac{(-\sigma^2)^r n!}{2^r r!(n-2r)!}X^{n-2r} \tag{5.3.18.4}$$

многочлены Эрмита, отмасштабированные так, чтобы быть унитарными и ортогональными относительно $\mathcal{N}(0,\sigma^2)$.

5.3.19. (Предполагаемая связь собственных ковекторов $\boldsymbol{u}_j$ с многочленами Эрмита.) Напомним, что собственные ковектора $\boldsymbol{u}_j$ образуют двойственный базис и потому связаны с $\boldsymbol{v}_k$ соотношениями ортогональности $\boldsymbol{u}_j^T\boldsymbol{v}_k = \delta_{jk}$, см. (5.3.12.3). Если представить себе, что координаты $\boldsymbol{u}_j$ также приблизительно задаются значениями некоторой гладкой функции $u_j(t)$ в точках $-m/2$, …, $m/2$, то соотношения ортогональности могут быть записаны как

$$\begin{aligned}\delta_{jk} &= \sum_{t=-m/2}^{m/2} \frac{(-1)^k}{k!}p_m^{(k)}(t)u_j(t) \approx \int_{\mathbb{R}} \frac{(-1)^k}{k!}p_m^{(k)}(t)u_j(t)\,dt \\ &\approx \frac{1}{k!\sqrt{2\pi}}\Big(\frac{m}{12}\Big)^{-k-1/2}\int_{\mathbb{R}} h_k^{[m/12]}(t)u_j(t)e^{-6t^2/m}dt\end{aligned} \tag{5.3.19.1}$$

согласно (5.3.18.3), т.е. функции $u_j(t)$ должны быть приблизительно ортогональны функциям $h_k^{[m/12]}(t)e^{-6t^2/m}$ относительно меры Лебега, или, что одно и тоже, многочленам Эрмита $h_k^{[m/12]}(t)$ относительно $\mathcal{N}(0,m/12)$. Хорошо известно, что для обычных многочленов Эрмита $(h_j(t),h_k(t))_{L^2(d\gamma)} = \delta_{jk}k!$, где γ — стандартная гауссова мера на $\mathbb{R}$, откуда

$$\int_{\mathbb{R}} h_j^{[\sigma^2]}(t)h_k^{[\sigma^2]}(t)\rho_\sigma(t)\,dt = \delta_{jk}k!\sigma^{2k} \tag{5.3.19.2}$$

для масштабированных многочленов Эрмита $h_n^{[\sigma^2]}(t)$ из (5.3.18.4). Применяя это равенство для $\sigma^2 = m/12$ и сравнивая его с (5.3.19.1), мы получаем следующую формулу, предположительно верную для компонент $\boldsymbol{u}_j$ с маленькими индексами j:

$$(\boldsymbol{u}_j)_t \approx h_j^{[m/12]}(t) \quad \text{для } t = -m/2,\, -m/2+1,\, \ldots,\, m/2 \tag{5.3.19.3}$$

Иначе говоря, мы вправе ожидать, что $(\boldsymbol{u}_j)_t$ ведут себя приблизительно как значения в точках $-m/2, \ldots, m/2$ некоторого унитарного многочлена степени j, близкого к многочлену Эрмита $h_j^{[m/12]}(t)$.

5.3.20. (Альтернативная формула для ковекторов $\boldsymbol{u}_j$.) Обратимся теперь к более тщательному изучению ковекторов $\boldsymbol{u}_j$, заданных несколько непонятной формулой (5.3.9.2) с числами Стирлинга:

$$(\boldsymbol{u}_{m-1-j})_s = \binom{m-1}{j}^{-1} \sum_{\ell=j}^{m-1} (-1)^{m-\ell-1} \begin{bmatrix} m-1 \\ \ell \end{bmatrix} \binom{\ell}{j} \Big(\frac{m}{2} + s - 1\Big)^{\ell-j} \tag{5.3.20.1}$$

при $0 \le j \le m-2$. Заметим, что это выражение — унитарный многочлен степени $m-1-j$ по s (благодаря нормировочному множителю $\binom{m-1}{j}^{-1}$). Обозначим через

$$\tilde{\boldsymbol{u}}_j := \binom{m-1}{j} \boldsymbol{u}_j \quad \text{при } 0 \le j \le m-1 \tag{5.3.20.2}$$

версию собственных ковекторов без нормировочного множителя. Переобозначим s буквой X, домножим получившееся выражение на Y^j и просуммируем по всем j:

$$\begin{aligned} \sum_{j=0}^{m-1} (\tilde{\boldsymbol{u}}_{m-1-j})_X Y^j &= \sum_{j=0}^{m-1} \sum_{\ell=j}^{m-1} (-1)^{m-\ell-1} \begin{bmatrix} m-1 \\ \ell \end{bmatrix} \binom{\ell}{j} (X + m/2 - 1)^{\ell-j} Y^j \\ &= \sum_{\ell=0}^{m-1} (-1)^{m-\ell-1} \begin{bmatrix} m-1 \\ \ell \end{bmatrix} (X + Y + m/2 - 1)^{\ell} \\ &= (X + Y + m/2 - 1)^{\underline{m-1}} \end{aligned} \tag{5.3.20.3}$$

Отсюда мы заключаем, что

$$(\tilde{\boldsymbol{u}}_{m-1-j})_s = [Y^j](Y + s + m/2 - 1)^{\underline{m-1}} = [Y^j] \prod_{k=s-m/2+1}^{s+m/2-1} (Y + k) \tag{5.3.20.4}$$

и

$$(\boldsymbol{u}_{m-1-j})_s = \binom{m-1}{j}^{-1} \cdot [Y^j] \prod_{k=s-m/2+1}^{s+m/2-1} (Y+k) \tag{5.3.20.5}$$

Из этой формулы сразу следует, например, что

$$\boldsymbol{u}_{m-1} = \tilde{\boldsymbol{u}}_{m-1} = (m-1)!\big(\boldsymbol{e}_{m/2} + (-1)^{m-1}\boldsymbol{e}_{-m/2}\big) \tag{5.3.20.6}$$

где $\boldsymbol{e}_s$ — стандартные базисные вектора, откуда ввиду (5.3.14.8)

$$\boldsymbol{u}'_{m-1} = \boldsymbol{u}_{m-1} = (m-1)!\big(\boldsymbol{e}_{m/2} + (-1)^{m-1}\boldsymbol{e}_{-m/2}\big) \tag{5.3.20.7}$$

$$\boldsymbol{u}'_{m} = \tilde{\boldsymbol{u}}_{m} - \frac{\boldsymbol{u}_{m-1}}{2} = \frac{(m-1)!}{2}\big(\boldsymbol{e}_{m/2} + (-1)^{m}\boldsymbol{e}_{-m/2}\big) \tag{5.3.20.8}$$

Кроме того, (5.3.20.5) при $j = m-1$ даёт

$$(\boldsymbol{u}_0)_s = (\tilde{\boldsymbol{u}}_0)_s = 1 \tag{5.3.20.9}$$

Впрочем, последнее равенство непосредственно следует и из (5.3.20.1).

5.3.21. (Эффективный алгоритм для вычисления всех ковекторов.) Заметим, что (5.3.20.4) может использоваться для очень эффективного вычисления всех $\tilde{\boldsymbol{u}}_j$ или $\boldsymbol{u}_j$, т.е. матрицы U. В самом деле, пусть $Q_s(Y) := \prod_{k=s-m/2+1}^{s+m/2-1}(Y+k)$. Вычислим $Q_{-m/2}(Y) = \prod_{k=1}^{m-1}(Y-k)$ непосредственно, и затем будем последовательно вычислять

$$Q_s(Y) = \frac{(Y+s+m/2-1)Q_{s-1}(Y)}{Y+s-m/2} \tag{5.3.21.1}$$

при $s = -m/2+1, \ldots, m/2$, запоминая каждый раз коэффициенты $Q_s(Y)$ в s-ую строку матрицы U согласно (5.3.20.5). Недостающий собственный вектор $\tilde{\boldsymbol{u}}_m$ или $\boldsymbol{u}'_m$ может быть затем вычислен по (5.3.9.4) или (5.3.20.8). Отметим, что также верно, что $Q_s(Y) = Q_{s-1}(Y+1)$, но это гораздо менее эффективно с вычислительной точки зрения, поскольку требует $O(m^2)$ операций вместо $O(m)$. Кроме того, рекуррентное соотношение (5.3.21.1) очень похоже на то, что используется при вычислении матрицы Кравчука: там надо начинать с многочлена $(1-Y)^m$ и на каждом шаге умножать его на $(1+Y)/(1-Y)$.

5.3.22. (Универсальные многочлены $L_n(m,X)$, задающие компоненты собственных ковекторов.) Отметим, что компоненты $\tilde{\boldsymbol{u}}_j$ при $0\le j\le m-1$ — это значения некоторого многочлена $\tilde L_j(m,X)$ степени $j-1$ относительно X в точках $-m/2$, $-m/2+1$, …, $m/2$, а компоненты $\boldsymbol{u}_j$ — это значения $L_j(m,X):=\binom{m-1}{j}^{-1}\tilde L_j(m,X)$. Посмотрим на эти многочлены повнимательнее, особенно при маленьких j. Согласно (5.3.20.1),

$$\tilde L_n(m,X)=\sum_{\ell=0}^{n}(-1)^\ell\begin{bmatrix}m-1\\ m-1-\ell\end{bmatrix}\binom{m-1-\ell}{n-\ell}\Big(X+\frac{m}{2}-1\Big)^{n-\ell}\tag{5.3.22.1}$$

и равенство (5.3.20.3) приобретает вид

$$\sum_{n=0}^{m-1}\tilde L_n(m,X)Y^{m-1-n}=(X+Y+m/2-1)^{\underline{m-1}}\tag{5.3.22.2}$$

Отметим, что (5.3.22.1) определяет $\tilde L_n(m,X)$ как многочлен от X для всех $m\in\mathbb N$ и $n\in\mathbb N_0$, даже если $n\ge m$, однако $\tilde L_n(m,X)=0$ при $n\ge m$, потому что тогда в каждом слагаемом формулы (5.3.22.1) обращается в ноль либо число Стирлинга $\left[\begin{smallmatrix}m-1\\ m-1-\ell\end{smallmatrix}\right]$ (при $\ell\ge m$), либо биномиальный коэффициент $\binom{m-1-\ell}{n-\ell}$ (при $\ell\le m-1$).

Хорошо известно, что $\left[\begin{smallmatrix}x\\ x-\ell\end{smallmatrix}\right]$ при фиксированном ℓ является многочленом степени 2ℓ относительно x. Например, $\left[\begin{smallmatrix}x\\ x\end{smallmatrix}\right]=1$, $\left[\begin{smallmatrix}x\\ x-1\end{smallmatrix}\right]=\binom{x}{2}=x(x-1)/2$, $\left[\begin{smallmatrix}x\\ x-2\end{smallmatrix}\right]=\frac{3x-1}{4}\binom{x}{3}$, $\left[\begin{smallmatrix}x\\ x-3\end{smallmatrix}\right]=\binom{x}{2}\binom{x}{4}$. Более общо,

$$\begin{bmatrix}x\\ x-n\end{bmatrix}=\sum_{k=0}^{n}\left\langle\!\!\left\langle{n\atop k}\right\rangle\!\!\right\rangle\binom{x+k}{2n}\tag{5.3.22.3}$$

где $\left\langle\!\!\left\langle{n\atop k}\right\rangle\!\!\right\rangle$ — эйлеровы числа второго порядка: $\left\langle\!\!\left\langle{0\atop k}\right\rangle\!\!\right\rangle=\delta_{k0}$, $\left\langle\!\!\left\langle{n\atop k}\right\rangle\!\!\right\rangle=(k+1)\left\langle\!\!\left\langle{n-1\atop k}\right\rangle\!\!\right\rangle+(2n-k-1)\left\langle\!\!\left\langle{n-1\atop k-1}\right\rangle\!\!\right\rangle$ при $n\ge1$. Отметим, что формула (5.3.22.3) выполнена для всех целых $x\ge0$, даже если $x<n$; в этом случае она даёт правильное (нулевое) значение.

Воспользуемся (5.3.22.3) для $x=m-1$. Тогда из (5.3.22.1) немедленно следует, что $\tilde L_n(m,X)$ — многочлен с рациональными коэффициентами от двух переменных m и X. Более того, $\tilde L_n(m,X)$ обращается в ноль при подстановке целых значений $m=1$, 2, …, n, и потому многочлен $\tilde L_n(m,X)$ делится на $\binom{m-1}{n}$. Отсюда следует, что $L_n(m,X)=$

$\binom{m-1}{n}^{-1}\tilde{L}_n(m,X)$ — также многочлен от m и X (а не всего лишь унитарный многочлен степени n от X, коэффициенты которого — рациональные функции от m), и мы можем вычислить $\tilde{L}_n(m,X)$, а затем и $L_n(m,X)$ при маленьких n, например, с помощью системы компьютерной алгебры:

$$L_0(m,X) = 1 \tag{5.3.22.4}$$
$$L_1(m,X) = X \tag{5.3.22.5}$$
$$L_2(m,X) = X^2 - \frac{m}{12} \tag{5.3.22.6}$$
$$L_3(m,X) = X^3 - \frac{m}{4}X \tag{5.3.22.7}$$
$$L_4(m,X) = X^4 - \frac{m}{2}X^2 + \frac{m(5m+2)}{240} \tag{5.3.22.8}$$
$$L_5(m,X) = X^5 - \frac{5m}{6}X^3 + \frac{m(5m+2)}{48}X \tag{5.3.22.9}$$
$$L_6(m,X) = X^6 - \frac{5m}{4}X^4 + \frac{m(5m+2)}{16}X^2 - \frac{m(35m^2+42m+16)}{4032} \tag{5.3.22.10}$$
$$L_7(m,X) = X^7 - \frac{7m}{4}X^5 + \frac{7m(5m+2)}{48}X^3 - \frac{m(35m^2+42m+16)}{576}X \tag{5.3.22.11}$$

Видно, что многочлены $L_n(m,X)$ представляют из себя некоторые дискретные приближения к многочленам Эрмита $h_n^{[m/12]}(X)$, наподобие многочленов Кравчука (ортогональных многочленов относительно биномиального распределения).

5.3.23. (Многочлены $L_n(m,X)$ задают $\boldsymbol{u}_n$ при $0 \le n < m$, но не $\tilde{\boldsymbol{u}}_m$.) Таким образом, координаты собственных ковекторов $\boldsymbol{u}_j$ при $0 \le j \le m-1$ действительно являются значениями некоторых универсальных многочленов степени j, в полном согласии с **5.3.19**:

$$(\boldsymbol{u}_j)_s = L_j(m,s) \quad \text{при } 0 \le j \le m-1,\ s = -m/2,\ -m/2+1,\ \ldots,\ m/2 \tag{5.3.23.1}$$

Отметим во избежание недоразумений, что оставшийся «вспомогательный» собственный ковектор $\tilde{\boldsymbol{u}}_m$ или его симметричная версия $\boldsymbol{u}'_m$ *не* задаются значениями $L_m(m,X)$. Впрочем, эти два ковектора обладают только одной или двумя ненулевыми компонентами и легко находятся по формуле (5.3.9.4) или (5.3.20.8).

5.3.24. (Многочлены $L_n(m,X)$ не являются последовательностью ортогональных многочленов.) Отметим, что многочлены $L_n(m,X)$ не являются ортогональными ни с каким весом, какое бы $m\in\mathbb{R}$ не было бы зафиксировано. В этом плане они принципиально отличаются от многочленов Кравчука. В самом деле, $\bigl(L_4(m,X)\bmod L_3(m,X)\bigr)\bmod L_2(m,X)=L_4(m,X)-XL_3(m,X)+\frac{m}{4}L_2(m,X)=\frac{m}{120}\neq 0$ при $m\neq 0$, а для ортогональных многочленов должен был бы получиться ноль. При $m=0$ многочлен $L_2(0,X)=X^2$ также не может быть ортогонален $L_0(0,X)=1$ ни с каким весом.

5.3.25. (Чётность многочленов $L_n(m,X)$.) Проверим, что многочлены $L_n(m,X)$ четны (относительно X) при чётном n, и нечётны при нечётном:

$$L_n(m,-X)=(-1)^nL_n(m,X)\quad\text{для всех } n\in\mathbb{N}_0 \tag{5.3.25.1}$$

Достаточно проверить это тождество многочленов для бесконечного количества конкретных значений m, например, для всех целых $m\geq n+1$. В этом случае достаточно проверить его для $\tilde{L}_n(m,X)$. Однако многочлен

$$\sum_{n=0}^{m-1}\tilde{L}_n(m,X)Y^{m-n-1}=\prod_{t=-m/2+1}^{m/2-1}(X+Y+t) \tag{5.3.25.2}$$

из (5.3.22.2) умножается на $(-1)^{m-1}$ при одновременной замене $X\leftrightarrow -X$ и $Y\leftrightarrow -Y$. Отсюда получаем желаемое равенство $\tilde{L}_n(m,-X)=(-1)^n\tilde{L}_n(m,X)$ для всех целых $m\geq n+1$.

5.3.26. («Формула Родригеса» для $L_n(m,X)$.) Отметим, что согласно формуле Тейлора для многочленов тождество (5.3.25.2) равносильно

$$\tilde{L}_n(m,X)=\frac{1}{(m-n-1)!}\left(\frac{d}{dX}\right)^{m-n-1}\prod_{t=-m/2+1}^{m/2-1}(X+t) \tag{5.3.26.1}$$

или, поскольку $L_n(m,X)=\tilde{L}_n(m,X)/\binom{m-1}{n}$,

$$L_n(m,X)=\frac{n!}{(m-1)!}\left(\frac{d}{dX}\right)^{m-n-1}\prod_{t=-m/2+1}^{m/2-1}(X+t) \tag{5.3.26.2}$$

для всех целых $m\geq n+1$.

5.3.27. (Многочлены $L_n(m,X)$ образуют последовательность Аппеля.) Отметим, что из (5.3.26.2) немедленно следует, что *унитарные многочлены* $L_n(m,X)$ *образуют последовательность Аппеля:*

$$L'_n(m,X) = \frac{d}{dX}L_n(m,X) = nL_{n-1}(m,X) \tag{5.3.27.1}$$

В этом плане многочлены $L_n(m,X)$ лучше приближают многочлены Эрмита, чем многочлены Кравчука, поскольку многочлены Эрмита тоже образуют последовательность Аппеля. Более того, из этих двух свойств многочленов Эрмита — ортогональности и того, что они образуют последовательность Аппеля — невозможно сохранить более одного, поскольку последовательность многочленов Эрмита — единственная (с точностью до аффинной замены переменной) последовательность многочленов, являющаяся одновременно ортогональной (относительно какой-нибудь положительной меры) и последовательностью Аппеля.

5.3.28. (Центральные разности многочленов $L_n(m,X)$.) Помимо (5.3.27.1), многочлены $L_n(m,X)$ удовлетворяют его дискретному аналогу с центральными конечными разностями вместо производной:

$$\nabla L_n(m,X) = L_n\Big(m,X+\frac{1}{2}\Big) - L_n\Big(m,X-\frac{1}{2}\Big) = nL_{n-1}(m-1,X) \tag{5.3.28.1}$$

Достаточно проверить это тождество после подстановки всевозможных целых $m \geq n+1$. В этом случае оно равносильно

$$\tilde{L}_n\Big(m,X+\frac{1}{2}\Big) - \tilde{L}_n\Big(m,X-\frac{1}{2}\Big) = (m-1)\tilde{L}_{n-1}(m-1,X) \tag{5.3.28.2}$$

что немедленно проверяется с помощью производящей функции (5.3.25.2).

5.3.29. (Многочлены $L_n(m,X)$ при $m=0$.) Докажем, что

$$L_n(0,X) = X^n \tag{5.3.29.1}$$

или, что равносильно,

$$\tilde{L}_n(0,X) = \binom{-1}{n}X^n = (-X)^n \tag{5.3.29.2}$$

Для этого применим формулу (5.3.22.1) при $m=0$. Возникающие в ней числа Стирлинга $\left[\begin{smallmatrix}-1\\-\ell-1\end{smallmatrix}\right]$ должны быть определены по формуле (5.3.22.3),

которая даёт

$$\begin{bmatrix} -1 \\ -n-1 \end{bmatrix} = \sum_{k=0}^{n} \left\langle\!\left\langle {n \atop k} \right\rangle\!\right\rangle \binom{k-1}{2n} = \left\langle\!\left\langle {n \atop 0} \right\rangle\!\right\rangle \binom{-1}{2n} = \left\langle\!\left\langle {n \atop 0} \right\rangle\!\right\rangle = 1 \tag{5.3.29.3}$$

Подстановка этих значений в (5.3.22.1) даёт

$$\tilde{L}_n(0, X) = \sum_{\ell=0}^{n} (-1)^\ell \binom{-1-\ell}{n-\ell} (X-1)^{n-\ell} \tag{5.3.29.4}$$

Поскольку $\binom{-1-\ell}{n-\ell} = (-1)^{n-\ell}\binom{n}{\ell}$, эта сумма равна $(-1)^n X^n$ по биному.

5.3.30. (Свободные члены $L_n(m, X)$.) Пусть

$$a_r(m) := L_{2r}(m, 0) \tag{5.3.30.1}$$

— многочлены от m, являющиеся свободными членами многочленов L_n с чётными индексами. Например,

$$a_0(m) = 1 \tag{5.3.30.2}$$

$$a_1(m) = -\frac{m}{12} \tag{5.3.30.3}$$

$$a_2(m) = \frac{m(5m+2)}{240} \tag{5.3.30.4}$$

$$a_3(m) = -\frac{m(35m^2+42m+16)}{4032} \tag{5.3.30.5}$$

$$a_4(m) = \frac{m(5m+4)(35m^2+56m+36)}{34560} \tag{5.3.30.6}$$

$$a_5(m) = -\frac{m(385m^4+1540m^3+2684m^2+2288m+768)}{101376} \tag{5.3.30.7}$$

Поскольку $L_{2r+1}(m, 0) = 0$ согласно (5.3.25.1), и L_n образуют последовательность Аппеля (см. **5.3.27**), мы можем восстановить все L_n по последовательности их свободных членов $\{a_r(m)\}$:

$$L_n(m, X) = \sum_{\nu=0}^{\lfloor n/2 \rfloor} \binom{n}{2\nu} a_\nu(m) X^{n-2\nu} \tag{5.3.30.8}$$

Таким образом, если мы научимся вычислять многочлены $a_r(m)$, мы тем самым определим все $L_n(m, X)$.

5.3.31. (Рекуррентная формула для $a_r(m)$.) Мы можем вывести рекуррентную формулу для многочленов $a_r(m)$ следующим образом. Подставим $X = 0$, $n = 2r + 1$ в формулу (5.3.28.1):

$$L_{2r+1}\Big(m, \frac{1}{2}\Big) - L_{2r+1}\Big(m, \frac{1}{2}\Big) = (2r+1)L_{2r}(m-1, 0) = (2r+1)a_r(m-1) \tag{5.3.31.1}$$

Теперь выразим левую сторону этого равенства с помощью (5.3.30.8):

$$\begin{aligned}(2r+1)a_r(m-1) &= \sum_{\nu=0}^{r} \binom{2r+1}{2\nu} a_\nu(m) 4^{\nu-r} \\ &= (2r+1)a_r(m) + \sum_{\nu=0}^{r-1} \binom{2r+1}{2\nu} 4^{\nu-r} a_\nu(m)\end{aligned} \tag{5.3.31.2}$$

откуда получаем разностное уравнение

$$a_r(m) - a_r(m-1) = -\frac{1}{2r+1} \sum_{\nu=0}^{r-1} \binom{2r+1}{2\nu} 4^{\nu-r} a_\nu(m) \tag{5.3.31.3}$$

позволяющее найти $a_r(m)$, если уже известны все $\{a_\nu(m)\}_{0\le\nu\le r-1}$, поскольку $a_r(0) = 0$ при $r \ge 1$ по (5.3.29.1).

5.3.32. (Степень и старший коэффициент $a_r(m)$.) Отметим, что из (5.3.31.3) и $a_0(m) = 1$ по индукции немедленно следует, что $a_r(m)$ — многочлен степени (в точности) r. Более того, мы можем вычислить его старший коэффициент:

$$\begin{aligned}[m^r]a_r(m) &= \frac{1}{r}[m^{r-1}]\big(a_r(m) - a_r(m-1)\big) \\ &= -\frac{1}{r(2r+1)}\binom{2r+1}{2r-2} 4^{-1}[m^{r-1}]a_{r-1}(m)\end{aligned} \tag{5.3.32.1}$$

откуда

$$[m^r]a_r(m) = -\frac{2r-1}{12}[m^{r-1}]a_{r-1}(m) \tag{5.3.32.2}$$

и

$$a_r(m) = (-1)^r \frac{(2r-1)!!}{12^r} m^r + O(m^{r-1}) \tag{5.3.32.3}$$

Для сравнения, свободные члены многочленов Эрмита $h_n^{[m/12]}(X)$ задаются выражениями $h_{2r}^{[m/12]}(0) = (2r-1)!! \cdot (-m/12)^r$, т.е. если оставить от $a_r(m)$ только члены при старшей степени m^r, многочлены $L_n(m, X)$ превращаются в многочлены Эрмита $h_n^{[m/12]}(X)$, что уточняет приближенное равенство (5.3.19.3).

5.4 Расширенные цепи Маркова

Заметим, что мы могли бы не ограничивать с самого начала множество состояний цепи Маркова из предыдущего пункта $m+1$ числами $-m/2$, $-m/2+1$, …, $m/2$, а могли бы разрешить более широкое множество состояний $-M/2 \dots M/2$, где $M \geq m$ — произвольное целое число той же чётности, что и m. Посмотрим, что изменяется при таком расширении множества состояний.

5.4.1. (Распространение результатов на более широкое множество состояний.) В действительности наша цепь Маркова очень быстро убегает на бесконечность из состояний s с $|s| > m/2$, так что новые состояния привносят в матрицу A_m огромную нильпотентную часть, но никак не меняют её ненулевые собственные числа и их кратность.

Чтобы убедиться в этом, зафиксируем какое-нибудь $M \geq m$ как выше, и пусть $\tilde{A}_m$ — матрица переходов цепи Маркова с расширенным множеством состояний. Таким образом, $\tilde{A}_m$ — это квадратная матрица порядка $M+1$, содержащая A_m в качестве подматрицы. Элементы $(\tilde{A}_m^n)_{ts}$ вычисляются с помощью ровно той же формулы (5.3.7.1), которая, однако, теперь применима, только если $N = 2^n \geq M$ (а не m). Поэтому мы можем определить матрицы $\boldsymbol{C}_j^{(m)}$, $0 \leq j \leq m-1$, и вектора $\boldsymbol{v}_j$, $\tilde{\boldsymbol{v}}_m$, $\boldsymbol{u}_j$, $\tilde{\boldsymbol{u}}_m$ ровно так же, как и раньше, но с более широкими множествами индексов координат. Единственное существенное отличие — формулу (5.3.9.4) для $\tilde{\boldsymbol{u}}_m$ теперь надо записывать как

$$(\tilde{\boldsymbol{u}}_m)_s = (\boldsymbol{u}_{m-1})_s \cdot [s \geq m/2] = (\boldsymbol{u}_{m-1})_s \cdot [s > 0] \tag{5.4.1.1}$$

где мы больше не можем пользоваться тем, что $[s \geq m/2] = \delta_{s,m/2}$, но зато воспользовались тем, что $(\boldsymbol{u}_{m-1})_s = 0$ при $0 \leq s < m/2$, и потому условие $s \geq m/2$ можно заменить на более красивое условие $s > 0$. Симметричная версия $\boldsymbol{u}'_m = \tilde{\boldsymbol{u}}_m - \boldsymbol{u}_{m-1}/2$ будет поэтому теперь равна

$$(\boldsymbol{u}_m)_s = (\boldsymbol{u}'_m)_s = \frac{\operatorname{sgn}(s)}{2}(\boldsymbol{u}_{m-1})_s \tag{5.4.1.2}$$

У ковекторов $\boldsymbol{u}_j$ возникает много новых ненулевых компонент, которые могут быть вычислены с помощью тех же универсальных многочленов

$L_j(m, X)$ из **5.3.22**:

$$(\boldsymbol{u}_j)_s = L_j(m, s) \quad \text{для } 0 \le j \le m-1 \text{ и всех } s \equiv \frac{m}{2} \pmod 1 \tag{5.4.1.3}$$

$$(\boldsymbol{u}_m)_s = \frac{\operatorname{sgn}(s)}{2} L_{m-1}(m, s) \quad \text{для тех же } s \tag{5.4.1.4}$$

Напротив, все новые компоненты векторов $\boldsymbol{v}_j$ окажутся равны нулю, а старые по-прежнему будут задаваться формулами (5.3.13.3), которые, впрочем, дают правильное нулевое значение и для новых компонент. Разложения (5.3.9.6) и (5.3.9.5) для $\tilde{A}_m^n$ и для расширенных матриц $\boldsymbol{C}_j^{(m)}$ будут по-прежнему верны, и применение **5.3.10** позволяет нам заключить, что собственные числа $\tilde{A}_m$ — это 0 и 2^{-j-1} при $0 \le j \le m-1$; кратность 2^{-j-1} равна рангу матрицы $\boldsymbol{C}_j^{(m)}$, который не мог уменьшиться от расширения, и потому кратность собственного числа 2^{-j-1} равна единице при $0 \le j < m-1$ и двойке при $j = m-1$. Соответственно, кратность нулевого собственного числа равна $M - m$. Кроме того, (расширенные) вектора $\boldsymbol{v}_j$ по-прежнему являются собственными векторами, а $\boldsymbol{u}_j$ — собственными ковекторами матрицы $\tilde{A}_m$.

5.4.2. (Свойства проектора $\boldsymbol{P}$ на сумму собственных пространств, соответствующих ненулевым собственным числам.) Пожалуй, самое большое отличие расширенной матрицы $\tilde{A}_m$ от исходной A_m заключается в том, что проектор

$$\boldsymbol{P} = \sum_{j=0}^{m-1} \boldsymbol{C}_j^{(m)} = \sum_{j=0}^{m} \boldsymbol{v}_j \boldsymbol{u}_j^T \tag{5.4.2.1}$$

более не равен единичной матрице, а является идемпотентной матрицей ($\boldsymbol{P}^2 = \boldsymbol{P}$) ранга $m+1$, задающей проекцию расширенного векторного пространства $\tilde{V} = k^{M+1}$ на сумму собственных подпространств, отвечающих ненулевым собственным числам матрицы $\tilde{A}_m$. Впрочем, соотношения ортогональности $\boldsymbol{u}_j^T \boldsymbol{v}_k = \delta_{jk}$ сохраняются, например, потому что вектора $\boldsymbol{v}_j$ в разложении (5.4.2.1) не перестали быть линейно независимыми после добавления дополнительных компонент, равно как и $\boldsymbol{u}_j$, откуда и из идемпотентности $\boldsymbol{P}$ следует ортогональность (либо можно заметить, что новые компоненты $\boldsymbol{v}_k$ равны нулю, а старые компоненты $\boldsymbol{u}_j$ не изменились, так что не изменилось и скалярное произведение). Заметим, что все новые компоненты собственных векторов $\boldsymbol{v}_j$ равны нулю; поэтому образ $\boldsymbol{P}$, т.е. сумма всех собственных подпространств, соответствующих ненулевым собственным числам, есть $(m+1)$-мерное пространство всех

векторов, все новые компоненты которых равны нулю. Таким образом, проектор $\boldsymbol{P}$ «сжимает» расширенные вектора $\boldsymbol{y}$ до векторов, все ненулевые компоненты которых имеют индексы $|s| \le m/2$. Это «сжатие» можно описать следующим образом: компоненты $\boldsymbol{y}_t$ большого вектора $\boldsymbol{y}$ суммируются с полиномиальными весами $(\boldsymbol{u}_j)_t = L_j(m,t)$ (за исключением $j = m-1$, для которого формула не полиномиальная), и затем полученные коэффициенты c_j домножаются на кусочно-полиномиальные функции $(-1)^j p_m^{(j)}(t)/j!$ и суммируются. Итоговая кусочно-полиномиальная функция вычисляется в точках $t = -m/2,\ -m/2+1,\ \ldots,\ m/2$, что и даёт все ненулевые компоненты проекции $\boldsymbol{Py}$.

Отметим, что при желании мы могли бы вычислить значения «сжатой» кусочно-полиномиальной функции $\sum_j (-1)^j c_j p^{(j)}(t)/j!$ не только в целых или полуцелых точках t, но и в произвольных вещественных точках t, желательно принадлежащих отрезку $[-m/2, m/2]$ (иначе получится ноль). В каком-то смысле образ $\boldsymbol{Py}$ — это именно эта кусочно-полиномиальная функция, а не только её значения на дискретном множестве точек $-m/2,\ -m/2+1,\ \ldots,\ m/2$.

5.4.3. (Свойства транспонированного проектора $\boldsymbol{P}^T$.) Транспонированный проектор

$$\boldsymbol{P}^T = \sum_{j=0}^{m} \boldsymbol{u}_j \boldsymbol{v}_j^T \tag{5.4.3.1}$$

напротив, распространяет вектора $\boldsymbol{y}$, у которых известны только координаты $\boldsymbol{y}_s$ с $|s| \le m/2$ (остальные координаты несущественны, поскольку не влияют на $\boldsymbol{v}_j^T \boldsymbol{y}$) на всё множество индексов s, причём так, что $\boldsymbol{P}^T\boldsymbol{y}$ совпадают с $\boldsymbol{y}$ в «старых» индексах $|s| \le m/2$. Иначе говоря, $\boldsymbol{P}^T$ осуществляет нечто вроде полиномиальной экстраполяции функции, изначально заданной только в целых или полуцелых точках $-m/2,\ -m/2+1,\ \ldots,\ m/2$, на более широкое множество аргументов: подбирается некоторый многочлен — линейная комбинация многочленов $L_j(m,X)$ при $0 \le j \le m-1$, принимающий данные значения y_s в исходных точках, а затем этот многочлен вычисляется во всех нужных s. Это описание совсем верно, если существует многочлен $P(X)$ степени не больше $m-1$, принимающий заданные значения $P(s) = \boldsymbol{y}_s$ при $s = -m/2,\ \ldots,\ m/2$. В противном случае следует учесть тот факт, что компоненты $(\boldsymbol{u}_m)_s$ — это не совсем значения многочлена, и экстраполяция осуществляется с помощью суммы некоторого многочлена степени не больше $m-1$ и некоторого

кратного функции $s \mapsto \mathrm{sgn}(s) L_{m-1}(m, s)$. Интересно, что получающаяся полиномиальная или кусочно-полиномиальная функция также может быть вычислена в произвольных вещественных точках s, и потому можно считать, что $\boldsymbol{P}^T \boldsymbol{y}$ — это на самом деле эта кусочно-полиномиальная функция, а не вектор её значений в точках $-M/2$, $-M/2+1$, ..., $M/2$.

5.4.4. (Внутреннее представление векторов или функций, ассоциированное с $\boldsymbol{P}$. Проекция на векторное пространство, порождённое первыми несколькими многочленами Эрмита.) Отметим, что внутреннее представление векторов $\boldsymbol{y}$ или функций, ассоциированное с проектором $\boldsymbol{P}$ (например, лежащих в его образе) — это последовательность $\{c_j\}_{0 \le j \le m}$ коэффициентов их разложения по собственным векторам $\boldsymbol{v}_j$, т.е. $c_j = \boldsymbol{u}_j^T \boldsymbol{y}$; тогда $\boldsymbol{P}\boldsymbol{y} = \sum_j c_j \boldsymbol{v}_j$. Аналогичное утверждение верно и для $\boldsymbol{P}^T$, с заменой $\boldsymbol{v}_j \leftrightarrow \boldsymbol{u}_j$. Мы видели, что первые несколько собственных векторов или ковекторов — это по существу приближения к многочленам Эрмита $h_j^{[m/12]}(t)$, которые после перемасштабирования можно заменить просто на $h_j(t)$. Здесь лучше рассматривать только $0 \le j \le m-1$; при $j = m$ эта закономерность несколько сбивается. Таким образом, $\boldsymbol{P}$ или (скорее) $\boldsymbol{P}^T$ по существу соответствуют ортогональной проекции всех функций из $L^2(\mathbb{R}, d\gamma)$ на конечномерное подпространство, порождённое первыми m многочленами Эрмита.

Список литературы

[GHS] Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics **129**, Cambridge University Press, 1997.

[DS] Alexander S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics **156**, Springer-Verlag, 1995.

[TG] Nicolas Bourbaki, *Topologie Générale*, Hermann, Paris, 1971 (chap. 1–4), 1974 (chap. 5–10).

[OP] Gabor Szegö, *Orthogonal polynomials*, AMS, 1978.

[TT] Peter Johnstone, *On a topological topos*, Proc. London Math. Soc. (3) **38** (1979) pp. 237–271.

[CH1] Th. Coquand, G. Huet, *Constructions: A higher order proof system for mechanizing mathematics.* In EUROCAL'85, volume **203**, Linz, 1985. Springer-Verlag.

[CH2] Th. Coquand, G. Huet, *The Calculus of Constructions,* Information and Computation, **76** (2/3), 1988.

[CP] Th. Coquand, C. Paulin-Mohring. *Inductively defined types,* in P. Martin-Löf and G. Mints, editors, Proceedings of Colog'88, **417**. Springer-Verlag, 1990.

[SP] Thierry de la Rue, *Espaces de Lebesgue*, Séminaire de Probabilités XXVII, Lecture Notes in Mathematics **1557**, Springer, 1993, pp. 15–21.

[DF] Bruno de Finetti, *La prévision: ses lois logiques, ses sources subjectives*, Annales de l'institut Henri Poincaré, **7** (1937) no. 1, pp. 1–68.

[AB1] Дуров Н.В., *Обзор подходов построения абсолютной геометрии*, Препринт ПОМИ 8 (2022).

[AB3] Дуров Н.В., *Индуктивные и коиндуктивные конструкции в математике*, Препринт ПОМИ 9 (2022).

[AB4] Дуров Н.В., *Примеры эффективных систем счисления*, Препринт ПОМИ 10 (2022).

[AB5] Дуров Н.В., *Вероятностные свойства избыточных систем счисления*, Препринт ПОМИ 11 (2022).

[AB6] Дуров Н.В., *Регулярные пространства и регулярные отображения*, Препринт ПОМИ 12 (2022).