Закономерности в словах стохастических контекстно-свободных языков, порожденных грамматиками с двумя классами нетерминальных символов. Вопросы экономного кодирования

А.Е.Борисов

Часть I

Введение

I.1 Постановка задачи и общая характеристика работы

Вопросы оптимального кодирования сообщений (слов), порождаемых некоторым вероятностным источником, возникают в задачах, связанных с передачей и хранением информации. Под оптимальным кодированием, как правило, понимают кодирование, которое дает максимальную степень сжатия информации (отношение длины исходного слова к длине его кода). Возможная степень сжатия информации определяется вероятностными и структурными свойствами источника информации. Вероятностные свойства учитывают частоты символов или фрагментов в исходном слове, а структурные (синтаксические) - правила построения слов (например, в словах русского языка не встречается сочетаний типа БББ, ЩЩ, слова не начинаются с мягкого знака).

Хорошо известны и широко используются на практике алгоритмы побуквенного кодирования, которые учитывают только частоты букв. Наиболее известными являются алгоритм Хаффмана [31], Фано [22], Шеннона [21], алгоритм арифметического кодирования [29], [30], которые применяются при отсутствии априорных знаний о структуре кодируемых слов (сообщений). Существуют также динамические версии этих алгоритмов, которые в процессе работы обновляют таблицу частот символов. Такие алгоритмы строят вероятностную модель языка в процессе получения на вход новых символов.

Хорошо известны также словарные методы сжатия, основанные на учете часто повторяемых фрагментов в кодируемом тексте. Здесь следует выделить алгоритмы Лемпеля-Зива LZ77 и LZ78 [27],[28], и их многочисленные модификации.

Задача кодирования, учитывающего синтаксические свойства сообщений, была впервые рассмотрена К.Шенноном в [21]. Им был рассмотрен вероятностный источник сообщений с конечным числом состояний. Шеннон доказал, что в случае эргодичности источника все сообщения большой длины N разбиваются на два класса: множество сообщений, вероятность которых стремится к 0 при $N \to \infty$, и множество оставшихся сообщений с приблизительно одинаковыми вероятностями и буквенным составом. Такой источник фактически соответствует неразложимому марковскому процессу с конечным числом состояний. При построении алгоритмов кодирования Шенноном рассматривались в основном вероятностные свойства слов, хотя введенное им блочное кодирование косвенно учитывает и структурные свойства.

Вопросы кодирования слов с учетом структурных свойств (синтаксиса) рассматривались А.А.Марковым в [13], [14], [15]. Он поставил задачу кодирования не на всем множестве слов алфавита, а на некотором подмножестве слов, описываемом синтаксическими правилами. Для описания синтаксиса рассматривались в основном регулярные источники. Марков показал, что учет синтаксиса регулярных языков позволяет увеличить степень сжатия и уменьшить вычислительную сложность алгоритмов кодирования [15].

Ближайшим обобщением класса языков, порожденных регулярными источниками, являются языки, порожденные стохастическими контекстно-свободными (КС-) грамматиками. Неразложимые согласованные стохастические КС-грамматики рассматривались Л.П.Жильцовой в работах [9], [10], [11]. В этих работах были получены следующие результаты. В докритическом случае (перронов корень матрицы первых моментов меньше единицы) для слов большой длины стохастического КС-языка, порождаемого такой грамматикой, установлены свойства, аналогичные свойствам слов, генерируемых эргодическим конечным источником [21].

Кроме того, для множества деревьев вывода высоты t для слов языка при $t \to \infty$ были найдены математические ожидания числа применений каждого правила грамматики на фиксированном ярусе дерева вывода и во всем дереве вывода для критического (перронов корень матрицы первых моментов равен единице) и докритического случаев. В обоих случаях была найдена асимптотическая формула для энтропии множества слов, имеющих деревья вывода фиксированной высоты. Также была вычислена стоимость оптимального кодирования и построен алгоритм асимптотически оптимального кодирования для обоих случаев, имеющий полиномиальную временную сложность.

В данной работе аналогичные исследования проводятся для грамматики с разложимой матрицей первых моментов и двумя классами нетерминальных символов. Рассматривается докритический и критический случай. Для такой грамматики установлены закономерности в деревьях вывода фиксированной высоты t для слов языка при $t \to \infty$, то есть найдены математические ожидания числа применений правил грамматики на каждом ярусе дерева вывода и во всем дереве, а также найдена суммарная вероятность таких слов. Особое внимание уделяется случаю кратного перронова корня, поскольку именно тогда появляются существенные отличия от неразложимого случая. С использованием этих закономерностей найдена асимптотическая формула для энтропии множества слов, имеющих деревья вывода фиксированной высоты.

Полученные результаты использованы для получения нижней оценки стоимости двоичного кодирования слов языка, порожденного рассматриваемой грамматикой. Задача оптимального кодирования рассматривалась в том же виде, что и в [11]. Под оптимальным кодированием понималось взаимно-однозначное кодирование множества всех слов языка, которое позволяет хорошо сжимать длинные слова (имеющие деревья вывода большой высоты). Была найдена стоимость оптимального кодирования для рассматриваемого языка. Доказано, что алгоритм блочного кодирования, использованный Жильцовой в [11] для неразложимого случая, является асимптотически оптимальным и для рассматриваемой в данной работе грамматики. Используемый алгоритм кодирования использует состав правил в выводе слова, таким образом учитывая синтаксические свойства.

Также были изучены общие вопросы кодирования стохастических языков, получена нижняя оценка для средней длины кода в зависимости от энтропии, которая является более сильной, чем полученная в [11], при больших значениях энтропии. Эта оценка была использована для получения нижней оценки стоимости кодирования стохастического КС-языка.

получены

I.2 Основные определения и обозначения

Здесь изложена упрощенная система определений и понятий, необходимая для формулировки основных результатов. Будем использовать определения КС-грамматики и стохастического КС-языка, приведенные в [1], [19].

Языком над алфавитом $B = \{b_1, \ldots, b_n\}$ называется подмножество $L \subseteq B^*$. Стохастический язык $\mathcal{L}$ определяется как пара (L, P), где L - язык, а P - распределение вероятностей на множестве его слов, причем вероятность каждого слова языка больше нуля.

Стохастическая порождающая контекстно-свободная грамматика - это система $G = \langle V_T, V_N, R, s\rangle$, где V_T и V_N - множества терминальных и нетерминальных символов соответственно (в дальнейшем называемых *терминалами* и *нетерминалами*), $|V_N| = k$, $s \in V_N$ - *аксиома*, R - конечное множество правил, $R = \bigcup\limits_{i=1}^{k} R_i$, где $R_i = \{r_{i1}, \ldots, r_{in_i}\}$, и каждое правило в R_i имеет вид

$$r_{ij} : A_i \overset{p_{ij}}{\to} \beta_{ij},\ j = 1, \ldots, n_i, \text{ где } A_i \in V_N, \beta_{ij} \in (V_T \cup V_N)^*,$$

а p_{ij} - вероятность применения правила r_{ij}, удовлетворяющая условиям $0 < p_{ij} \leq 1,\ \sum\limits_{j=1}^{n_i} p_{ij} = 1$.

Применение правила к слову состоит в замене в этом слове нетерминала, стоящего в левой части правила, на правую часть применяемого правила.

Будем говорить, что слово α *непосредственно выводимо* из слова β, если $\alpha = \alpha_1 A_i \alpha_2$, $\beta = \alpha_1 \beta_{ij} \alpha_2$ для некоторых слов $\alpha_1, \alpha_2 \in (V_T \cup V_N)^*$, и грамматика содержит правило $r_{ij} : A_i \overset{p_{ij}}{\to} \beta_{ij}$. Рефлексивное транзитивное замыкание отношения непосредственной выводимости назовем *отношением выводимости.* За L_G обозначим множество слов в терминальном алфавите, выводимых из аксиомы s. Под *выводом слова* будем понимать последовательность правил грамматики, с помощью которой данное слово выводится из аксиомы. *Левый вывод* - это вывод, в котором каждое правило применяется к самому левому по порядку нетерминалу в слове.

Пусть $\alpha \in L_G$. Каждому левому выводу слова α соответствует *дерево вывода*, корень которого помечен аксиомой, а вершины - терминальными и нетерминальными символами [1]. При применении правила $A_i \xrightarrow{p_{ij}} h_1 h_2 \dots h_n$ в вершине a, помеченной нетерминалом A_i, добавляются n дуг от a к вершинам следующего яруса, которые помечаются слева направо символами $h_1, h_2, \dots, h_n,\ h_i \in V_T \cup V_N$. Все листья дерева помечены терминальными символами, при этом само слово получается обходом листьев дерева слева направо. Высотой дерева вывода называется наибольшая длина пути от корня до листа. Ярусы в дереве мы будем нумеровать с единицы, т.е. ярус, на котором располагается корень дерева, имеет номер один, следующий - два и т.д.

Если α - некоторое слово, которое имеет левый вывод $\omega = r_{i_1 j_1} \dots r_{i_k j_k}$, а d - соответствующее ему дерево вывода, то обозначим $p(d) = p_{i_1 j_1} \dots p_{i_k j_k}$. Пусть $D(L_G)$ - множество всех деревьев вывода для слов из L_G.

Стохастическая КС-грамматика называется *согласованной*, если

$$\lim_{N \to \infty} \sum_{d \in D(L_G), |d| < N} p(d) = 1,$$

где через $|d|$ обозначается высота дерева d. В дальнейшем будем рассматривать только согласованные грамматики. Величину $p(d)$ будем называть вероятностью дерева вывода d.

Стохастическим КС-языком, соответствующим стохастической КС-грамматике G, называется множество слов L_G с индуцированным на нем распределением вероятностей. Вероятность слова определяется как суммарная вероятность всех различных деревьев вывода для этого слова.

Пусть k - число нетерминалов в рассматриваемой грамматике. *Первые моменты* $a^i_j,\ i, j = 1, \dots, k$ грамматики определяются как $a^i_j = \sum\limits_{l=1}^{n_i} p_{il} s^{il}_j$, где s^{il}_j - число нетерминалов A_j в правой части правила r_{il}. Матрица $A = (a^i_j)$ называется *матрицей первых моментов. Вторые моменты* определяются как

$$b^i_{jm} = \sum_{l=1}^{n_i} p_{il} s^{il}_j (s^{il}_m - \delta^m_j),$$

где δ^m_j - символ Кронекера ($\delta^i_i = 1,\ \delta^i_j = 0$ при $i \neq j$).

Будем говорить, что нетерминал A_j следует за нетерминалом A_i, или что A_i предшествует A_j (и обозначать $A_i \to A_j$), если из A_i выводимо хотя бы одно слово, содержащее нетерминал A_j. Грамматика называется *неразложимой*, если для любых двух различных нетерминалов A_i и A_j верно $A_i \to A_j$. В противном случае она называется *разложимой. Классом* нетерминалов назовем максимальное по включению подмножество $K \in V_N$, такое, что $A_i \to A_j$ для любых $A_i, A_j \in K$.

Будем говорить, что класс K_1 предшествует классу K_2 (и обозначать $K_1 \prec K_2$), если для любых $A_1 \in K_1,\ A_2 \in K_2$ верно $A_1 \to A_2$. Очевидно, что множество нетерминалов V_N является объединением конечного числа непересекающихся классов. другому составляющих

Понятие неразложимости можно пояснить на языке графов. Построим ориентированный граф, в котором вершины соответствуют нетерминалам грамматики, и от i-й вершины к j-й проводится ребро, если существует правило вида $A_i \to \alpha_1 A_j \alpha_2,\ \alpha_1, \alpha_2 \in (V_T \cup V_N)^*$. Тогда неразложимой грамматике соответствует сильно связный граф, то есть такой, в котором существует ориентированный путь из любой вершины в любую.

Любой стохастической КС-грамматике можно очевидным образом поставить в соответствие дискретный ветвящийся процесс [16], причем нетерминалу грамматики будет соответствовать тип частицы в ветвящемся процессе. Классу нетерминалов грамматики будет соответствовать класс типов частиц в ветвящемся процессе.

Матрица или вектор X называется неотрицательной (обозначается $X \geq 0$), если все ее элементы неотрицательны, и положительной (обозначается $X > 0$), если все ее элементы больше нуля. Квадратная неотрицательная матрица X размера $k \times k$ называется разложимой, если множество $S = (1, 2, \ldots, k)$ можно разбить на два подмножества S_1, S_2 так, чтобы $X_{ji} = 0$ для любых $i \in S_1, j \in S_2$. В противном случае матрица X называется неразложимой. В [16] показано, что неразложимость ветвящегося процесса (грамматики) эквивалентна неразложимости матрицы первых моментов. Известно, что если неотрицательная матрица X неразложима и непериодична [8], то существует такое n, что $X^n > 0$. Очевидно, матрица первых моментов грамматики всегда неотрицательна.

По теореме Фробениуса [8] для матрицы первых моментов A существует максимальный по модулю действительный неотрицательный собственный корень (*перронов корень*), причем в случае $A > 0$ он простой, т.е. его кратность равна 1. Будем его обозначать в дальнейшем через r.

Пусть G - стохастическая КС-грамматика. Через L_i обозначим язык, порожденный грамматикой, которая получена заменой аксиомы исходной грамматики на нетерминал A_i. Через D_i обозначим множество деревьев вывода слов из L_i, через $D_i^{\leq t}$ - множество деревьев вывода высоты не более t для слов из L_i, а множество деревьев из $D_i^{\leq t}$ высоты ровно t обозначим через D_i^t. Для $d \in D_i^t$ обозначим через $p_t(d)$ условную вероятность дерева d. Она равна $p(d)/P(D_i^t)$, где $P(D_i^t)$ - общая вероятность деревьев вывода из D_i^t. Обозначим через $Q_i(t)$ вероятность множества деревьев из D_i, имеющих высоту больше чем t. Эту величину назовем вероятностью продолжения по аналогии с теорией ветвящихся процессов. Очевидно, что $P(D_i^t) = Q_i(t-1) - Q_i(t)$. Будем далее иногда для краткости обозначать $P_i(t) = P(D_i^t)$. Для исходной грамматики G будем полагать, что аксиомой является нетерминал A_1. В дальнейшем в обозначениях $L_i,\ D_i^t$ индекс i будем опускать при $i = 1$. Будем использовать также векторные обозначения $Q(t) = (Q_1(t), \ldots, Q_k(t)),\ P(t) = (P_1(t), \ldots, P_k(t))$.

В работе рассматривается разложимая согласованная грамматика с двумя классами нетерминалов K_1, K_2, причем один из них предшествует другому. Для определенности положим $K_1 \prec K_2$. Согласованность для грамматики означает, что соответствующий ей ветвящийся процесс вырождается, что, как показано в [16], имеет место либо при $r < 1$, либо если $r = 1$ и отсутствуют бесполезные (не участвующие в порождении слов языка) нетерминалы. Выполнение одного из этих двух условий мы и будем предполагать в дальнейшем. Считаем, что аксиома s принадлежит K_1. Обозначим $k_1 = |K_1|$, $k_2 = |K_2|$, $k_1 + k_2 = k$. Будем считать, что нетерминалы из K_1 имеют номера $1, \ldots, k_1$, из K_2 - соответственно $k_1 + 1, \ldots, k$. Для такой грамматики матрица первых моментов A имеет блочный вид:

$$A = \begin{pmatrix} A^{(1)} & A^{(2)} \\ 0 & A^{(3)} \end{pmatrix}, \tag{I.2.1}$$

где $A^{(1)}$ - матрица размером $k_1 \times k_1$, $A^{(3)}$ - матрица размером $k_2 \times k_2$, а $A^{(2)}$ - матрица размером $k_1 \times k_2$. Дополнительно предположим, что матрицы $A^{(1)}$, $A^{(2)}$, $A^{(3)}$ положительны. Тогда по теореме Перрона [8] для каждой из матриц $A^{(1)}, A^{(3)}$ существует простой максимальный по модулю действительный положительный собственный корень (перронов корень).

Обозначим за v', v'' левые, за u', u'' правые собственные вектора матриц $A^{(1)}$ и $A^{(3)}$, соответствующие их перроновым корням r' и r'' при нормировке $v'u' = v''u'' = 1$, $\sum_i v'_i = \sum_i v''_i = 1$ (левый собственный вектор матрицы всегда считаем вектором-строкой). Левый и правый собственные вектора для перронова корня r матрицы A при условии нормировки $vu = 1$, $\sum_i v_i = 1$ будем обозначать через v, u.

КС-грамматика называется КС-грамматикой с *однозначным выводом*, если каждому слову в терминальном алфавите, выводимому из аксиомы, соответствует единственное дерево вывода. При рассмотрении вопросов кодирования слов языка всегда будем рассматривать грамматику с однозначным выводом.

Опишем процедуру укрупнения грамматики с однозначным выводом, введенную в [11]. Пусть α - слово из $(V_T \cup V_N)^*$, выводимое из нетерминала A_i, и $d(\alpha)$ - его дерево вывода. α. Обозначим через M_i^n множество слов в алфавите $V_N \cup V_T$, выводимых из A_i и имеющих деревья вывода высоты не более n, в которых нетерминалами могут быть помечены листья только последнего (n-го) яруса. Укрупнение грамматики состоит в замене множества правил R на $R(n) = \cup_{i=1}^{k} R_i(n)$, где

$$R_i(n) = \{r'_{ij} : A_i \xrightarrow{p'_{ij}} \beta'_{ij}, j = 1, \ldots, n'_{ij},\ \beta'_{ij} \in M_i^n\},$$

где n'_{ij} - число слов в M_i^n. Вероятность правила $p'_{ij} = p(\beta'_{ij})$ равна вероятности соответствующего слова из M_i^n в исходной грамматике. Очевидно, что $L_G = L_{G(n)}$, и, если G - согласованная грамматика с однозначным выводом, то $G(n)$ - также согласованная грамматика с однозначным выводом. Данная процедура может применяться и к грамматике с неоднозначным выводом; в этом случае вероятность правила в грамматике $G(n)$ равна сумме вероятностей деревьев вывода для его правой части. В [16] показано, что матрица первых моментов для грамматики $G(n)$ равна A^n, где A - матрица первых моментов грамматики G.

Оказывается, что требование непериодичности матриц $A^{(1)}, A^{(3)}$ и условий $A^{(1)}$, $A^{(2)}$, $A^{(3)} > 0$ можно значительно ослабить. Допустим, что только матрица $A^{(3)}$ непериодична. Тогда легко показать, что с помощью укрупнения грамматики и отбрасывания бесполезных нетерминалов можно добиться непериодичности матрицы

$A^{(1)}$, а также положительности матриц $A^{(1)}$, $A^{(2)}$, $A^{(3)}$. Требование непериодичности матрицы $A^{(3)}$ здесь является существенным.

с хотя укрупнение

Рассмотрим стохастический КС-язык $\mathcal{L} = (L, P)$, порожденный грамматикой L с однозначным выводом. Под *кодированием* языка $\mathcal{L}$ будем понимать инъективное отображение $f : L \to \{0,1\}^+$. Образ слова при отображении f назовем его кодом. Множество всех кодирований языка $\mathcal{L}$ обозначим за $F(\mathcal{L})$.

Под *энтропией языка* $\mathcal{L}$ будем понимать величину

$$H(\mathcal{L}) = -\lim_{N\to\infty} \sum_{\alpha\in L, |\alpha|\le N} p(\alpha)\cdot\log p(\alpha),$$

где через $p(\alpha)$ обозначается вероятность слова α. Здесь и далее под log будем понимать логарифм по основанию 2. Если энтропия определена и конечна, то будем писать для краткости просто $H(\mathcal{L}) = -\sum\limits_{\alpha\in L} p(\alpha)\cdot\log p(\alpha)$.

За L^t обозначим множество слов языка L, деревья вывода которых имеют высоту t. Через $p_t(\alpha)$ обозначим условную вероятность слова α в множестве слов L^t, она равна $p(\alpha)/P(L^t)$. Через $\mathcal{L}^t = (L^t, p_t)$ обозначим множество слов L^t с определенным на нем выше распределением вероятностей. *Стоимостью кодирования* $f \in F(\mathcal{L})$ назовем величину

$$C(\mathcal{L}, f) = \lim_{t\to\infty} \frac{\sum_{\alpha\in L^t} p_t(\alpha)\cdot|f(\alpha)|}{\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha|},$$

если этот предел существует. Через $F_*(\mathcal{L})$ обозначим множество всех кодирований, для которых величина $C(\mathcal{L}, f)$ определена. Стоимостью оптимального кодирования назовем величину $C^*(\mathcal{L}) = \inf\limits_{f\in F_*(\mathcal{L})} C(\mathcal{L}, f)$. Этот подход к определению стоимости кодирования был рассмотрен Жильцовой в [11]. В некотором смысле такая постановка задачи эквивилентна задаче кодирования длинных слов, порожденных эргодическим источником, которая была рассмотрена Шенноном в [21].

Для произвольного стохастического языка $\mathcal{L}$ (не обязательно контекстно-свободного), предел математического ожидания длины кодового слова

$$M(\mathcal{L}, f) = \lim_{N\to\infty} \sum_{\alpha\in L, |\alpha|\le N} p(\alpha)\cdot|f(\alpha)|$$

назовем *средней длиной кода* для кодирования f. Здесь за $|x|$ обозначается длина слова x. Если эта величина существует и конечна, то будем писать $M(\mathcal{L}, f) = \sum\limits_{\alpha\in L} p(\alpha)\cdot|f(\alpha)|$. Значение $M^*(\mathcal{L}) = \inf\limits_{f\in F(L)} M(\mathcal{L}, f)$ будем называть *оптимальной средней длиной кода.*

I.3 Формулировка основных результатов

Диссертация состоит из введения и трех глав.

I. Во второй главе рассматриваются вопросы оптимального кодирования произвольного стохастического языка. Изучена связь между оптимальной средней длиной кода и энтропией для произвольного стохастического языка. В [11] доказано, что если энтропия $H(\mathcal{L})$ конечна, то оптимальная средняя длина кода $M^*(\mathcal{L})$ тоже конечна и удовлетворяет неравенствам

$$H(\mathcal{L})/2 \leq M^*(\mathcal{L}) < H(\mathcal{L}) + 1$$

В этой главе получена более точная нижняя оценка для $M^*(\mathcal{L})$, которая выполняется для больших значений энтропии, а именно, если энтропия стохастического языка $\mathcal{L} = (L, P)$ удовлетворяет неравенству $1 < H(\mathcal{L}) < \infty$, то справедливы следующие утверждения:
1) выполняется неравенство:

$$H(\mathcal{L}) \leq \left(1 + \log\left(\frac{M^*(\mathcal{L})}{M^*(\mathcal{L}) - 1}\right)\right) \cdot M^*(\mathcal{L}) + \log(M^*(\mathcal{L}) - 1);$$

2) для любого $H_* > 1$ существует такой стохастический язык $\mathcal{L}$, для которого $H(\mathcal{L}) = H_*$, а неравенство пункта 1) превращается в равенство (теорема II.2);
3) Отношение $M^*(\mathcal{L})/H(\mathcal{L})$ достигает нижней оценки 1/2 (теорема II.1) только при $H(\mathcal{L}) = H_* = 4$. При $H(\mathcal{L}) \to \infty$ и при $H(\mathcal{L}) \to 1$ отношение $M^*(\mathcal{L})/H(\mathcal{L})$ возрастает и стремится к верхней оценке 1 (теорема II.3).

II. В третьей главе рассматривается разложимая грамматика с двумя классами нетерминалов в докритическом случае (перронов корень $r < 1$). Сначала для соответствующего грамматике ветвящегося процесса устанавливается асимптотика вероятностей продолжения, которая для неразложимого случая получена в [16]. Установлено, что для случая перронова корня кратности один вероятности продолжения имеют такую же асимптотику, как и в неразложимом случае, а для перронова корня кратности два их асимптотика имеет другой вид, а именно:

$$Q_i(t) \sim bc_0 u'_i t r^t, i = 1, \ldots, k_1,$$

$$Q_i(t) \sim c_0 u''_{i-k_1} r^t, i = k_1 + 1, \ldots, k,$$

где r-перронов корень, u', u'' - правые собственные вектора матриц $A^{(1)}$, $A^{(3)}$ соответственно, $c_0 > 0$ - некоторая константа, и b определено формулой (III.2.3) (теорема (III.2.1)). В доказательствах использовалась техника из теории ветвящихся процессов, аналогичная примененной в [16] для неразложимых процессов.

С использованием этих результатов показано, что в случае простого перронова корня вероятности продолжения, вероятность деревьев вывода высоты t при $t \to \infty$, и математические ожидания числа применения правил на фиксированном ярусе дерева вывода и во всем дереве имеют такую же асимптотику, как и в неразложимом случае. Это позволило доказать, что в этом случае энтропия $H(\mathcal{L}^t)$ множества слов, имеющих деревья вывода высоты t, и стоимость оптимального кодирования имеют тот же вид, что и в неразложимом случае.

В случае кратного перронова корня математическое ожидания числа применений правила r_{ij} в дереве высоты t на ярусе τ линейно зависит от τ/t при $t, \tau \to \infty$, и при этом ограничено сверху константой:

$$M(q_{ij}(t,\tau)) \sim \frac{p_{ij}\cdot(t-\tau)}{t}\cdot\left(G_i' + \frac{v_i'S_{ij}'}{r}\right)$$

для $i \le k_1$,

$$M(q_{ij}(t,\tau)) \sim \frac{p_{ij}}{t}\cdot\left(\tau\cdot(v_{i-k_1}''S_{ij}''/r + G_i'') + (t-\tau)\cdot G_i'\right)$$

для $k_1 < i \le k$ (теорема (III.5.1)).

Доказано, что для правил, применяемых к нетерминалам из K_1, это величина убывает, а для правил, применяемых к нетерминалам из K_2, может и убывать, и возрастать (см. пример параграфа (III.7). При этом, как и в неразложимом случае, имеет место перераспределение частот правил, т.е. при $t \to \infty$ правила, порождающие больше нетерминальных символов, используются в выводе слов чаще, что, по-видимому, нужно для достижения большей высоты дерева вывода. Величины $M(q_{ij}(t,\tau))$ не меняются при замене аксиомы на любой нетерминал из класса K_1.

Также доказано, что среднее число правил, приходящихся на один ярус дерева вывода высоты t, стремится к константе при $t \to \infty$:

$$M(S_{ij}(t)/t) \to \omega_{ij},\ i = 1,\ldots,k,$$

где величины ω_{ij} определены формулами (III.5.6) (теорема (III.5.2)).

Дисперсия этой величины для случая кратного перронова корня $r < 1$ не стремится к 0 при $t \to \infty$ в отличие от неразложимого случая [9], а именно, справедливы соотношения

$$D(S_{ij}(t)/t) \to \left(\omega_{ij}^{(1)}\right)^2/12 \text{ при } i \le k_1,$$

$$D(S_{ij}(t)/t) \to \left(\omega_{ij}^{(2)} - \omega_{ij}^{(1)}\right)^2/12 \text{ при } i > k_1,$$

где величины $\omega_{ij}^{(l)}$ определены формулами (III.5.5) (теорема (III.6.1)).

Поэтому пропорциональный состав деревьев вывода по правилам грамматики и пропорциональный побуквенный состав слов соответствующего языка не имеет места.

Для нахождения математические ожидания числа применения правил грамматики на ярусе дерева вывода, среднего числа правил, приходящихся на один ярус дерева вывода, и его дисперсии был использован метод, примененный Жильцовой для неразложимого случая [9].

III. С помощью этих результатов получено, что энтропия $H(\mathcal{L}^t)$ множества слов, имеющих деревья вывода высоты t, асимптотически линейно зависит от t, как и в неразложимом случае:

$$H(\mathcal{L}^t) = t\cdot\left(\log r - \sum_{i=1}^{k}\sum_{j=1}^{n_i}\omega_{ij}\log p_{ij}\right) + o(t).$$

Здесь величины ω_{ij} определены формулами (III.5.6), а p_{ij} - вероятность применения правила r_{ij} (теорема (III.8.1)).

Из результатов главы 2 следует, что для оптимального кодирования такого множества слов отношение математического ожидания длины закодированного слова к $H(\mathcal{L}^t)$ стремится к 1 при $t \to \infty$. Это позволило получить нижнюю оценку $C^*(\mathcal{L})$ для стоимости оптимального (двоичного) кодирования слов языка, порождаемого рассматриваемой грамматикой:

$$C^*(\mathcal{L}) = \frac{\log r - \sum_{i,j} \omega_{ij} \log p_{ij}}{\sum_{i,j} l_{ij}\omega_{ij}},$$

где ω_{ij} определены формулами (III.5.6), p_{ij} - вероятность применения правила r_{ij}, а l_{ij} - число терминальных символов в правой части правила r_{ij} (теорема (III.8.2)).

Кроме того, показано, что эта оценка является асимптотически точной, то есть существует кодирование, стоимость которого сколь угодно близка к полученной оценке $C^*(\mathcal{L})$. Для этого использовалась схема кодирования, предложенная в [11], которая состоит в применении префиксного кодирования Шеннона для множества правил R_i с одинаковой левой частью A_i с учетом найденных математических ожиданий числа применений правил в деревьях вывода. Код слова получается конкатенацией кодов правил, примененных при его выводе. По аналогии с [11] доказывается, что стоимость такого кодирования стремится к $C^*(\mathcal{L})$ при описанном в этой работе укрупнении правил грамматики. Такое кодирование существенно использует и вероятностные и структурные свойства языка (вероятности применения и частоты правил, а также состав их правой части).

IV. В четвертой главе рассмотрена разложимая грамматика с двумя классами нетерминалов в критическом случае (перронов корень $r = 1$).

Сначала рассматривается случай кратного перронова корня. Для вероятностей продолжения соответствующего грамматике ветвящегося процесса аналогично неразложимому случаю [16] получаются рекуррентные соотношения. В результате их решения получено, что асимптотика вероятностей продолжения для нетерминалов первого класса и вероятность деревьев вывода высоты t в случае кратного перронова корня имеет другой вид по сравнению с неразложимым случаем. Доказан следующий результат.

При условии $B_1 B_2 > 0$, где B_1, B_2 определяются вторыми моментами грамматики и даны формулами (IV.1.1.14), при $t \to \infty$ верны следующие асимптотические равенства:

$$Q_i(t) \sim u_i' k_0 t^{-1/2},$$

$$P_i(t) \sim \frac{u_i' k_0}{2t^{3/2}},$$

где $k_0 = \sqrt{\frac{4b}{B_1 B_2}}$, а b введено формулой (III.2.3) и определяется по матрице первых моментов (теорема IV.1.1.2).

Таким образом, нетерминалы из класса K_1 в некотором смысле более долгоживущие - т.е. вероятности продолжения $Q_i(t)$ при $i \le k_1$ больше, чем $Q_i(t)$ при $i > k_1$.

Поскольку не удалось доказать пропорциональность состава нетерминалов на фиксированном ярусе дерева вывода, то нахождение асимптотик математических ожиданий числа применений правил на каждом ярусе дерева вывода с помощью метода, примененного в докритическом случае, вызывает трудности. Вместо этого сразу была найдена асимптотика математического ожидания $M(S_{ij}(t))$ числа применений правила r_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ в деревьях вывода высоты t при $t \to \infty$ (здесь и везде далее через $M(\xi), D(\xi)$ обозначаются соответственно математическое ожидание и дисперсия случайной величины ξ). Для этого были получены и решены рекуррентные соотношения для величин $M(S_{ij}(t))$. Доказано, что при $r' = r'' = 1$ и $B_1 B_2 > 0$ справедливы следующие асимптотические равенства:

$$M(S_{ij}(t)) \sim \frac{p_{ij} B_2 t}{4b},\ i \le k_1,$$

$$M(S_{ij}(t)) \sim p_{ij} B_2 t^2/4,\ i > k_1,$$

где b дано формулой (III.2.3), а p_{ij} - вероятность применения правила r_{ij} (теорема IV.1.2.1).

Из этой теоремы видно, что в случае кратного перронова корня $r = 1$ для разложимой грамматики с двумя классами нетерминалов количество правил, примененных к нетерминалам второго класса, имеет такую же ($O(t^2)$) асимптотику, как и в неразложимом случае, а количество правил, примененных к нетерминалам первого класса, значительно меньше ($O(t)$). Таким образом, между нетерминалами 1-го и 2-го класса соблюдается баланс: нетерминалы 1-го класса имеют большую вероятность продолжения, но их меньше в деревьях вывода. Как и в неразложимом критическом случае, здесь наблюдается сохранение первоначальных относительных частот применения в дереве вывода правил с одинаковым нетерминалом в левой части, т.е. $M(S_{ij}(t))/p_{ij}$ не зависит от j.

V. С помощью этих результатов найдена асимптотика величины $H(\mathcal{L}^t)$ для случая $r = r' = r'' = 1,\ B_1 B_2 > 0:$

$$H(\mathcal{L}^t) = \sum_{i>k_1} v''_{i-k_1} H(R_i) \cdot B_2 t^2 \cdot (1 + o(1))/4,$$

где $H(R_i) = -\sum_{j=1}^{n_j} p_{ij} \log p_{ij}$ - энтропия множества правил R_i (теорема IV.1.3.1).

Эта теорема показывает, что величина $H(\mathcal{L}^t)$ определяется только правилами, применяемыми ко второму нетерминалу, поскольку в этом случае нетерминалов из класса K_2 в дереве вывода много больше, чем нетерминалов из K_1 (но не совпадает с соответствующей величиной для грамматики, порожденной только нетерминалами класса K_2).

С использованием этой теоремы и результатов главы 2 найдена нижняя оценка стоимости оптимального кодирования $C^*(\mathcal{L})$ для $r = r' = r'' = 1,\ B_1 B_2 > 0:$

$$C^*(\mathcal{L}) = \sum_{i>k_1} v''_{i-k_1} H(R_i) / \sum_{i>k_1} v''_{i-k_1} L(R_i).$$

где $H(R_i)$ - энтропия множества правил R_i, а $L(R_i)$ - среднее число терминалов в правой части правил из R_i (теорема IV.1.3.2).

Таким образом, в случае кратного перронова корня нижняя оценка стоимости оптимального кодирования определяется свойствами правил, применяемых к нетерминалам из класса K_2, и совпадает с соответствующей величиной для неразложимой грамматики, порожденной нетерминалами класса K_2. Это вызвано тем, что правил, применяемых к нетерминалам из K_2, значительно больше, чем правил, применяемых к нетерминалам из K_1.

Как и в докритическом случае, показывается, что эта оценка является точной. Оказывается, что оптимальное префиксное кодирование всех слов языка по методу Шеннона является оптимальным и с точки зрения кодирования слов с деревьями вывод большой высоты. Кроме того, показано, что блочное кодирование, использованное в предшествующей главе, также является асимптотически оптимальным.

VI. Далее рассматривается случай простого перронова корня $r = 1$. Оказывается, что в этом случае, как и при $r < 1$, асимптотики для величин $Q_i(t), P_i(t)$, а также математических ожиданий числа применения правил на фиксированном ярусе τ дерева вывода из D_i^t и во всем дереве имеют такой же вид, как и в неразложимом случае. Доказательства проводятся аналогично неразложимому критическому случаю [10], [16].

Пусть $M(x_i(t,\tau))$ - математическое ожидание числа нетерминалов A_i на ярусе τ в деревьях вывода высоты t, а $M(q_{ij}(t,\tau))$ - математическое ожидание числа применения правил r_{ij} на ярусе τ в деревьях вывода высоты t. В случае $r' < r'' = 1$, когда $v_i = 0,\ i = 1, \dots, k_1$, для величин $M(x_i(t,\tau)),\ M(q_{ij}(t,\tau)),\ M(S_{ij}(t))$ получены следующие оценки для $i \le k_1$ при $t, \tau \to \infty$:

$$M(x_i(t,\tau)) < C_1 \cdot \left(t^2 (r')^\tau / (t-\tau)\right),$$

$$M(q_{ij}(t,\tau)) < C_2 \cdot \left(t^2 (r')^\tau / (t-\tau)\right).$$

для некоторых $C_1, C_2 > 0$, и

$$M(\bar{S}_{ij}^q(t)) \to p_{ij} z_q^i,$$

$$M(S_{ij}^q(t)) \to p_{ij} \cdot \left(S_{ij} + \sum_m G_m^q p_{ij} z_m^i\right) z_q^i / u_q,$$

где $S_{ij},\ G_m^q,\ z_q^i$ - вычисляемые по грамматике константы (теорема IV.2.3.2).

Таким образом, в случае $0 < r' < r'' = 1$ в дереве вывода на каждом ярусе нетерминалов из класса K_2 значительно больше, чем из K_1. Оценки для $H(\mathcal{L}^t)$ и $C^*(\mathcal{L})$ в этом случае имеют тот же вид, что и для неразложимой грамматики, порожденной только нетерминалами из K_2.

Выражения для энтропии $H(\mathcal{L}^t)$ множества слов $\mathcal{L}^t$ и нижней оценки стоимости оптимального кодирования $C^*(\mathcal{L})$ при $0 < r' \neq r'' = 1$ имеют такой же вид, как и в неразложимом критическом случае, причем схема кодирования, использованная для случая кратного перронова корня, является оптимальной и в этом случае.

Часть II

Оценка средней длины кода для стохастического языка

В этой главе мы рассмотрим связь нижней оценки средней длины кода с энтропией для произвольного стохастического языка. Рассмотрим стохастический язык $\mathcal{L} = (L, P)$. Пусть $f \in F(\mathcal{L})$ - его кодирование. Кодирование f^*, для которого средняя длина кода $M(\mathcal{L}, f^*) = M^*(\mathcal{L})$ минимальна, будем называть (только в этом параграфе) оптимальным кодированием стохастического языка. Очевидно, оптимальным является кодирование, при котором слова языка упорядочиваются в порядке невозрастания вероятностей и кодируются последовательно сначала двоичными словами длины 1, потом словами длины 2 и т.д. Иногда в выкладках будем для краткости обозначать $M^*(\mathcal{L})$ за M^*, $H(\mathcal{L})$ за H когда речь идет об одном и том же языке $\mathcal{L}$.

Жильцовой доказана следующая теорема.

Теорема II.1 *[11] Пусть $\mathcal{L} = (L, P)$ - стохастический язык, для которого энтропия $H(\mathcal{L})$ конечна. Тогда средняя длина оптимального кода также конечна и удовлетворяет неравенствам*

$$H(\mathcal{L})/2 \leq M^*(\mathcal{L}) < H(\mathcal{L}) + 1.$$

Получим несколько более точных нижних оценок для $M^*(\mathcal{L})$, которые выполняются для больших значений энтропии. Докажем следующую теорему.

Теорема II.2 *Пусть энтропия стохастического языка (конечного или бесконечного) $\mathcal{L} = (L, P)$ удовлетворяет неравенству $1 < H(\mathcal{L}) < \infty$, причем количество слов в L больше 2. Тогда:*
1) Справедливо неравенство $M^*(\mathcal{L}) > 1$;
2) Справедливо неравенство:

$$H(\mathcal{L}) \leq \left(1 + \log\Big(\frac{M^*(\mathcal{L})}{M^*(\mathcal{L}) - 1}\Big)\right) \cdot M^*(\mathcal{L}) + \log(M^*(\mathcal{L}) - 1);$$

3) для любого $H^* > 1$ существует (бесконечный) стохастический язык $\mathcal{L}$, для которого $H(\mathcal{L}) = H^*$, а неравенство пункта 2) превращается в равенство.

Доказательство. Упорядочим слова языка в порядке невозрастания вероятностей, так что $p_1 \geq p_2 \geq \ldots$, где p_i-вероятность i-го слова. В [11] показано, что при оптимальном кодировании длина кода i-го слова равна $\lfloor \log(i+1) \rfloor$, где $\lfloor x \rfloor$ - ближайшее целое снизу для x. Утверждение 1) следует из того, что $M^*(\mathcal{L}) = \sum_i p_i \lfloor \log(i+1) \rfloor \geq \sum_i p_i = 1$, и равенство достигается только если число слов в языке не превосходит двух. Докажем теперь утверждение 2).

Возьмем некоторое $a > 0$, и рассмотрим разность $H - (1 + a)M^*$. Пользуясь определениями энтропии и средней длины кода, получаем:

$$H - (1+a)M^* = \sum_{i>0} p_i \cdot \Big(\log(1/p_i) - (1+a)\lfloor \log(i+1) \rfloor\Big) = \sum_{i>0} p_i \cdot \log\big(2^{-(1+a)\lfloor \log(i+1)\rfloor}/p_i\big).$$

Применяя неравенство Йенсена для функции $\log(x)$, получаем, что

$$\sum_{i>0} p_i \cdot \log(2^{-(1+a)\lfloor \log(i+1)\rfloor}/p_i) \leq \log\left(\sum_{i>0} 2^{-(1+a)\lfloor \log(i+1)\rfloor}\right)$$

$$= \log\Big(2 \cdot 2^{-(1+a)} + 4 \cdot 2^{-2(1+a)} + \dots\Big) = \log\left(\sum_{i>0} 2^{-i \cdot a}\right) = -\log(2^a - 1).$$

Таким образом, $H \leq (1+a)M^* - \log(2^a - 1) = g(M^*, a)$. Легко проверить, что $g(M^*, a)$ выпукла вниз и при $M^* > 1$ достигает минимума в точке $a^* = \log(M^*/(M^* - 1))$. Подставляя $a = a^*$ в неравенство $H \leq g(M^*, a)$, получаем первую часть утверждения теоремы.

Для доказательства второй части утверждения заметим, что функция

$$g(M^*, a^*(M^*)) = g_0(M^*) = M^* \cdot (1 + \log(M^*/(M^* - 1))) + \log(M^* - 1)$$

монотонно возрастает и непрерывна при $M^* > 1$, так как $g_0'(M^*) > 0$ при $M^* > 1$. Заметим, что $g_0(M^*)$ можно записать в виде $g_0(M^*) = M^* + M^* \log(M^*) + (1 - M^*) \log(M^* - 1)$, поэтому

$$\lim_{M^* \to 1} g_0(M^*) = 1 + \lim_{M^* \to 1} (1 - M^*) \log(M^* - 1) = 1.$$

Ввиду монотонности и непрерывности $g_0(M^*)$ при $M^* > 1$ для любого $H_* > 1$ найдется $M^* > 1$, такое, что $H_* = g_0(M^*)$. Утверждение 2) доказано.

Докажем теперь утверждение 3). Для M^*, такого, что $H_* = g_0(M^*) = g(a^*, M^*)$, рассмотрим стохастический язык, в котором вероятность i-го слова $p_i = (2^{a^*} - 1) \cdot 2^{-(1+a^*)\lfloor \log(i+1)\rfloor}$. Нетрудно показать, что $\sum_{i>0} p_i = 1$, поэтому такое распределение вероятностей существует. Энтропия языка с таким распределением вероятностей определяется следующим образом:

$$H = -\sum_{i>0} p_i \log p_i = \sum_{k>0} 2^k \cdot (2^{a^*} - 1) \cdot 2^{-(1+a^*)k} \left((1 + a^*)k - \log(2^{a^*} - 1)\right)$$

$$= -\log(2^{a^*} - 1) + (1 + a^*) \cdot (2^{a^*} - 1) \cdot \sum_{k>0} k \cdot 2^{-a^* k} = (1 + a^*)2^{a^*}/(2^{a^*} - 1) - \log(2^{a^*} - 1)$$

$$= (1 + a^*)M^* - \log(2^{a^*} - 1).$$

Оптимальная средняя длина кода имеет вид:

$$M = \sum_{i>0} p_i l_i = (2^{a^*} - 1) \cdot \sum_{k>0} k \cdot 2^{-ka^*} = 2^{a^*}/(2^{a^*} - 1) = M^*,$$

где за l_i обозначена длина кода i-го слова при оптимальном кодировании. Таким образом, для рассмотренного языка неравенство пункта 2) леммы превращается в равенство, и теорема доказана.

Характер полученной оценки стоимости оптимального кодирования $M^*(\mathcal{L})$ в зависимости от $H(\mathcal{L})$ можно понять, посмотрев на график функции $H(x) = \Big(1 + \log(x/(x-1))\Big)x + \log(x-1)$. Она монотонно возрастает при $x > 1$, и $H(x) \to 1$ при $x \to 1$, поэтому в области $H > 1$ существует обратная функция $x(H) = M^*(H)$.

Функция $H(x)/x$ стремится к 1 при $x \to 1$ или $x \to \infty$, и достигает максимума $H(x)/x = 2$ при $x = 2$. Таким образом, минимум отношения стоимости кодирования и энтропии равен нижней оценке 1/2, полученной в теореме II.1. При $H \to \infty$ отношение стоимости кодирования к энтропии растет и сколь угодно приближается к верхней оценке 1.

Следствие II.1 *Для любого стохастического языка при $H = H(\mathcal{L}) > 2$ верно следующее неравенство (где $M^* = M^*(\mathcal{L})$):*

$$M^*/H \geq 1 - \frac{\log H}{H} - \frac{2\log e}{H-2}.$$

Доказательство. Используя неравенство $H/2 \leq M^* < H + 1$, получим, что:

$$M^* \geq \frac{H - \log(M^* - 1)}{1 + \log\big(M^*/(M^*-1)\big)} \geq \frac{H - \log H}{1 + \log\big(H/(H-2)\big)}.$$

Используя неравенства $\log(1+x) < x\log e$, $1/(1+x) > 1-x$ и $(1-x)(1-y) > 1-x-y$, справедливые при $x, y > 0$, имеем:

$$M^* \geq H \cdot \left(\frac{1 - \log H/H}{1 + 2\log e/(H-2)}\right) \geq H \cdot \left(1 - \frac{\log H}{H} - \frac{2\log e}{H-2}\right).$$

Следствие доказано.

Это следствие показывает, что $M^*(H)/H \to 1$ при $H \to \infty$, и позволяет получить следующий важный результат, который мы будем в дальнейшем использовать для получения нижней оценки стоимости кодирования стохастического языка.

Теорема II.3 *Если имеется последовательность стохастических языков $\mathcal{L}_t$, такая, что $H(\mathcal{L}_t) \to \infty$ при $t \to \infty$, то $\lim\limits_{t\to\infty} M^*(\mathcal{L}_t)/H(\mathcal{L}_t) = 1$.*

Доказательство. Пусть $H_t = H(\mathcal{L}_t)$, $M_t^* = M^*(\mathcal{L}_t)$. В силу следствия II.1 и теоремы II.2 имеем, что

$$1 + 1/H_t > M_t^*/H_t \geq 1 - \frac{\log H_t}{H_t} - \frac{2\log e}{H_t - 2}.$$

Переходя к пределу при $t \to \infty$, получаем:

$$1 \leq \lim_{t\to\infty} M_t^*/H_t \leq 1.$$

Часть III

Закономерности в деревьях вывода слов и оптимальное кодирование. Докритический случай

В этой части будем рассматривать случай, когда соответствующий грамматике ветвящийся процесс докритический, т.е. когда неразложимый ветвящийся процесс, порожденный каждым из классов K_1, K_2, докритический. Условие докритичности эквивалентно тому, что перроновы корни r', r'' матриц $A^{(1)},\ A^{(3)}$ меньше единицы. Таким образом, в дальнейшем будем полагать, что $0 < r' < 1$ и $0 < r'' < 1$. Отдельно будем рассматривать три случая : $r' > r'',\ r' < r'',\ r' = r''$.

III.1 Основные определения и обозначения

Введем некоторые обозначения и соглашения. Рассматривая вектор, будем считать, что мы имеем дело с вектором-столбцом, если противное не оговорено специально. Сумму элементов вектора-столбца или вектора-строки v будем обозначать через $\bar{v}$. В дальнейшем для вектора или матрицы X будем писать $X = c$ (соответственно $X \leq c,\ X \geq c,\ X < c,\ X > c$), где c - скаляр, если все компоненты вектора или матрицы X равны (соответственно меньше или равны, больше или равны, меньше или больше) c. Под записью $|X| < c$ будем понимать, что все компоненты вектора или матрицы X по модулю меньше скаляра c.

Кроме того, для двух векторов или матриц X_1, X_2 одинаковой размерности будем писать, что $X_1 = X_2$ ($X_1 \leq X_2,\ X_1 \geq X_2,\ X_1 < X_2,\ X_1 > X_2$), если все компоненты вектора или матрицы X_1 равны (соответственно меньше или равны, больше или равны, меньше или больше) компонентам X_2. Через $\bar{c}$ обозначим вектор, все компоненты которого равны константе - скаляру c. Записью $X/c,\ cX,\ X + c$, где X- вектор или матрица, будем обозначать соответственно деление, умножение всех элементов X на скаляр c или прибавление этого скаляра ко всем элементам X. За X^T обозначим транспонирование вектора или матрицы X. Через $\min(X), \max(X)$ будем обозначать соответственно минимальный(максимальный) элемент вектора (матрицы) X.

Пусть $x = (x_1, \ldots, x_n),\ \alpha = (\alpha_1, \ldots, \alpha_n)$. Будем обозначать

$$\alpha! = \alpha_1! \cdot \ldots \cdot \alpha_n!,\ x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}.$$

Через $v(m : n)$, где $v = (v_1, \ldots, v_m, \ldots, v_n, \ldots, v_k)$ - вектор, будем обозначать вектор $(v_m, \ldots, v_n)$. Через (v_1, v_2), где v_1, v_2 - вектора-столбцы, будем обозначать объединенный вектор-столбец $(v_1^T, v_2^T)^T$. Запись $x(t) \sim y(t)$ при $t \to \infty$ обозначает $x(t) = y(t) \cdot (1 + o(1))$.

Пусть $s=(s_1,\dots,s_k)$. Введем многомерные производящие функции [16]:

$$F_i(s)=F_i(s_1,\dots,s_k)=\sum_{j=1}^{n_i} p_{ij}\cdot s_1^{l_{ij}^1}\cdot\ldots\cdot s_k^{l_{ij}^k},\ i=1,\dots,k,$$

где l_{ij}^m - число вхождений нетерминального символа $A_m,\ m=1,\dots,k$ в правую часть правила r_{ij}. Тогда первые моменты грамматики определяются как $a_j^i=\frac{\partial F_i(s)}{\partial s_j}|_{s=1}$, и матрица $A=(a_j^i)$ называется матрицей первых моментов (очевидно, такое определение матрицы первых моментов эквивалентно более простому, приведенному в первой главе).

Пусть $\Xi=(\xi_1,\dots,\xi_k)$ - случайный вектор, $\alpha^*=(\alpha_1,\dots\alpha_k)$ - фиксированный вектор с целочисленными неотрицательными компонентами и $\alpha=\alpha_1+\ldots+\alpha_k$. Рассмотрим величину $\Xi^{\alpha^*}=\xi_1^{[\alpha_1]}\cdot\ldots\cdot\xi_k^{[\alpha_k]}$, где $\xi_i^{[\alpha_i]}=\xi_i(\xi_i-1)\dots(\xi_i-\alpha_i+1)$. Математическое ожидание $M\Xi^{\alpha^*}$ называется факториальным моментом вектора Ξ порядка α.

Обозначим через $x_j^i(t)$ число нетерминальных символов A_j в дереве вывода из D_i на ярусе t. Факториальный момент вектора $X^i(t)=(x_1^i(t),\dots,x_k^i(t))$ порядка α будем называть факториальным моментом грамматики порядка α и обозначать $m_\alpha^i(t)$. По аналогии с [16] введем обозначения $a_j^i(t)$ для факториальных моментов первого порядка вектора $X^i(t)$, причем $a_j^i(1)$ совпадают с введенными ранее величинами a_j^i. Для факториальных моментов второго порядка введем обозначения $b_{jj}^i(t)=Mx_j^i(t)(x_j^i(t)-1)$ и $b_{jk}^i(t)=Mx_j^i(t)x_k^i(t)$ $(j\neq k)$. Факториальные моменты 3-го порядка будем обозначать через $c_{jkl}^i(t)$. Как и для первых моментов, положим $b_{jk}^i=b_{jk}^i(1),\ c_{jkl}^i=c_{jkl}^i(1)$. Из [16] известно, что $b_{jk}^i=\frac{\partial^2 F_i(s)}{\partial s_j\partial s_k}|_{s=1},\ c_{jkl}^i=\frac{\partial^3 F_i(s)}{\partial s_j\partial s_k\partial s_l}|_{s=1}$.

Производящая функция и вероятности продолжения ветвящегося процесса [16], соответствующего грамматике, совпадают с соответствующими характеристиками грамматики. Это позволяет использовать результаты теории ветвящихся процессов для стохастических КС-грамматик.

III.2 Вероятности продолжения

В этом разделе будут получены асимптотические формулы для вероятностей продолжения $Q_i(t), i=1,\dots,k$. Нам понадобится следующая лемма.

Лемма III.2.1 *Пусть матрица A имеет вид (I.2.1). и ее перронов корень $r\le 1$. Тогда при $r'\neq r''$ максимальный собственный корень r для A простой и равен $\max\{r',r''\}$. Соответствующие ему левый и правый собственные вектора могут быть выбраны неотрицательными. Кроме того:*
1) При $r'>r''$ правый собственный вектор для A может быть выбран в виде $u=(u',0)$, а левый - в виде $v=(v',v^2)$, где $v^2>0$ - некоторый положительный вектор.
2) При $r'<r''$ правый собственный вектор для A может быть выбран в виде $u=(u^1,u'')$, а левый - в виде $v=(0,v'')$, где $u^1>0$ - некоторый положительный вектор.

Доказательство. Из равенства $|A-\lambda E| = |A^{(1)}-\lambda E|\cdot|A^{(3)}-\lambda E|$ вытекает, что любой собственный корень матрицы A является собственным корнем для $A^{(1)}$ или $A^{(3)}$, и наоборот. Отсюда следует простота r при $r' \neq r''$ и равенство $r = \max\{r', r''\}$.

Рассмотрим сначала случай $r' > r''$. Очевидно, вектор $u = (u', 0)$ является правым собственным вектором для r'. Вектор $v = (v^1, v^2)$ является левым собственным вектором матрицы A для перронова корня $r = r'$, если

$$v^1 A^{(2)} + v^2 A^{(3)} = r' v^2,\ \ v^1 A^{(1)} = r' v^1, \tag{III.2.1}$$

где $v^1 = (v^1_1, \ldots, v^1_{k_1}),\ v^2 = (v^2_1, \ldots, v^2_{k_2})$. Таким образом, $v^1 = v' > 0$. Легко показать, что

$$v^2 = v^1 A^{(2)}\cdot(r'E - A^{(3)})^{-1} = (r')^{-1} v^1 A^{(2)}\cdot\sum_{n=0}^{\infty}\Big((r')^{-1}A^{(3)}\Big)^n. \tag{III.2.2}$$

Действительно, последний ряд сходится, поскольку из [8] известно, что элементы матрицы $(A^{(3)})^n$ имеют вид $O\big((r'')^n\big)$, поэтому элементы матрицы $\Big((r')^{-1}A^{(3)}\Big)^n$ имеют вид $O\big((r''/r')^n\big)$. Прямой проверкой устанавливается, что матрица $E + \sum_{n=1}^{\infty}((r')^{-1}A^{(3)})^n$ является обратной к $E-(r')^{-1}A^{(3)}$. Следовательно, вектор $v = (v', v^2)$ удовлетворяет равенствам (III.2.1). Из положительности матриц $A^{(2)}, A^{(3)}$ и равенства (III.2.2) следует, что $v^2 > 0$. Доказательство утверждения леммы для случая $r' < r''$ проводится аналогично.

В дальнейшем в качестве собственных векторов для A будем рассматривать вектора со свойствами, установленными в лемме III.2.1. Асимптотика вероятностей продолжения описывается следующей теоремой.

Теорема III.2.1 *Для разложимой грамматики с двумя классами нетерминалов, матрица первых моментов которой имеет вид (I.2.1):*
1) при $r' \neq r''$ и $t \to \infty$ выполняется равенство

$$Q_i(t) = c_0 u_i r^t + o(r^t);$$

2) при $r = r' = r''$ и $t \to \infty$

$$Q_i(t) = b c_0 u'_i t r^t\cdot(1+o(1)), i = 1, \ldots, k_1,$$

$$Q_i(t) = c_0 u''_{i-k_1} r^t\cdot(1+o(1)), i = k_1+1, \ldots, k.$$

Здесь r', r'' - перроновы корни матриц $A^{(1)}$ и $A^{(3)}$ соответственно, $r = \max(r', r'') < 1$ - перронов корень матрицы первых моментов, u, v - соответствующие ему правый и левый собственные вектора, $c_0 > 0$ - некоторая константа, и b определено формулой

$$b = v' A^{(2)} u''/r. \tag{III.2.3}$$

Для доказательства этой теоремы нам понадобится несколько лемм, аналогичных леммам, доказанным в [16].

Лемма III.2.2 *Пусть матрица A имеет вид (I.2.1), r', r'' - перроновы корни матриц $A^{(1)}$ и $A^{(3)}$ соответственно, $r = \max(r', r'') \le 1$. Тогда:*
1) при $r' \ne r''$ верно асимптотическое равенство

$$A^t = uvr^t + O(r_1^t),$$

где $0 < r_1 < r$, $u \ge 0$ и $v \ge 0$;
2) при $r' = r'' = r$ верно асимптотическое равенство

$$A^t = \begin{pmatrix} u'v'r^t + O(r_1^t) & bu'v'' \cdot tr^t + o(tr^t) \\ 0 & u''v''r^t + O(r_1^t) \end{pmatrix},$$

где $r_1 < r$, и $b = v'A^{(2)}u''/r$.

Первая часть утверждения доказывается так же, как и для неразложимой матрицы (см. [8] или [16]), существование и неотрицательность собственных векторов для перронова корня при $r' \ne r''$ доказана в лемме III.2.1. Вторая часть утверждения следует из равенства

$$A^t = \begin{pmatrix} (A^{(1)})^t & \sum\limits_{i=0}^{t-1} (A^{(1)})^i A^{(2)} (A^{(3)})^{t-1-i} \\ 0 & (A^{(3)})^t \end{pmatrix}$$

и соотношений $(A^{(1)})^t = u'v'r^t + O(r_1^t)$, $(A^{(3)})^t = u''v''r^t + O(r_1^t)$, $0 < r_1 < r$.

Из этой леммы и равенства $A^t = (a_i^j)^t = (a_i^j(t))$, легко получить вид первых моментов грамматики.

Следствие III.2.1 *Пусть матрица A имеет вид (I.2.1) и ее перронов корень $r \le 1$. Тогда в обозначениях предыдущей леммы:*
1) при $r' \ne r''$

$$a_j^i(t) = u_i v_j r^t + O(r_1^t);$$

2) при $r = r' = r''$

$$a_j^i(t) = u_i' v_j' r^t + O(r_1^t), \ i, j \le k_1;$$

$$a_j^i(t) = bu_i' v_{j-k_1}'' tr^t + o(tr^t), \ i \le k_1, \ j > k_1;$$

$$a_j^i(t) = u_{i-k_1}'' v_{j-k_1}'' r^t + O(r_1^t), \ i, j > k_1,$$

где $0 \le r_1 < r$.

Лемма III.2.3 *Пусть матрица A имеет вид (I.2.1), r', r'' - перроновы корни матриц $A^{(1)}$ и $A^{(3)}$ соответственно, $r = \max(r', r'') \le 1$, и*

$$A_t = \begin{pmatrix} A_t^{(1)} & A_t^{(2)} \\ 0 & A_t^{(3)} \end{pmatrix}$$

- последовательность матриц, таких, что $A_t^{(1)}$, $A_t^{(2)}$, $A_t^{(3)}$ - матрицы размера $k_1 \times k_1$, $k_1 \times k_2$, $k_2 \times k_2$ соответственно, $0 \le A_t \le A$, и $A_t \to 0$ при $t \to \infty$. Пусть

$A_t^* = (A - A_1)(A - A_2)\dots(A - A_t)$. *Тогда для любого вектора* $x > 0$:
1) при $r' \neq r''$ *выполняется равенство*

$$\lim_{t\to\infty} \frac{A_t^* x}{v A_t^* x} = u; \qquad \text{(III.2.4)}$$

2) при $r' = r'' = r$ *верны равенства*

$$\lim_{t\to\infty} \frac{(A_t^* x)(1:k_1)}{v' \cdot (A_t^* x)(1:k_1)} = u',$$

$$\lim_{t\to\infty} \frac{(A_t^* x)(k_1+1:k)}{v'' \cdot (A_t^* x)(k_1+1:k)} = u''. \qquad \text{(III.2.5)}$$

Кроме того, в этом случае $\max((A_t^* x)(1:k_1))/\min((A_t^* x)(k_1+1:k)) \to 0$ *при* $t \to \infty$.

Схема доказательства этой леммы аналогична схеме доказательства соответствующего утверждения для неразложимой матрицы A, приведенного в [16]. Не уменьшая общности, можно считать, что $r = \max(r', r'') = 1$, домножив матрицы A, A_t на константу $1/r$ в случае, если это не так.

Рассмотрим случай $r' \neq r''$. Очевидно,

$$A_t^* x = (A - A_t)\dots(A - A_{n+1}) \cdot A_n^* x \text{ при } t > n.$$

Зададим произвольно малые величины $0 < \varepsilon_1, \varepsilon_2 < 1$, и выберем достаточно большое $l = l(\varepsilon_1)$ так, чтобы выполнялось неравенство $|A^l - uv| < \varepsilon_1$. Это возможно по утверждению (1) леммы III.2.2. Выберем $n = n(\varepsilon_2)$ так, чтобы $A_t < \varepsilon_2 A$ для всех $t > n$. Пусть $t = l + n$. Тогда верно неравенство

$$(1-\varepsilon_2)^{t-n} A^{t-n} x \le (A - A_t)\dots(A - A_{n+1}) x \le A^{t-n} x, \qquad \text{(III.2.6)}$$

так как $A - A_i \ge 0$.

Пусть $m_0 = \max_{d,i,j}(A_{ij}^d)$ (по лемме III.2.2 $m_0 < \infty$). Тогда

$$A^l - (1-\varepsilon_2)^l A^l \le \big(1 - (1-\varepsilon_2)^l\big) \cdot m_0. \qquad \text{(III.2.7)}$$

Обозначим $x^{(n)} = A_n^* x$. В силу неравенств (III.2.6), (III.2.7) и неравенства $|A^l - uv| < \varepsilon_1$ будет верно неравенство

$$\left|A_t^* x - uvx^{(n)}\right| \le \left|(A_t^* x - A^l x^{(n)}\right| + \left|A^l x^{(n)} - uvx^{(n)}\right| \le \left|\big(A^l - (1-\varepsilon_2)^l A^l\big) x^{(n)}\right| + \left|(uv - A^l) x^{(n)}\right|$$

$$< k \cdot \Big(\big(1 - (1-\varepsilon_2)^{l(\varepsilon_1)}\big) \cdot m_0 + \varepsilon_1\Big) \cdot \max x^{(n)}. \qquad \text{(III.2.8)}$$

Устремляя ε_1 к нулю, затем ε_2 к нулю, найдем n_0, l такие, что $|A_{n+l}^* x - uvx^{(n)}| < \varepsilon \max x^{(n)}$ для любого наперед заданного $\varepsilon > 0$ при $n > n_0$.

Обозначим

$$x'^{(n)} = x^{(n)}(1:k_1), x''^{(n)} = x^{(n)}(k_1+1:k), u^1 = u(1:k_1), u^2 = u(k_1+1:k).$$

Далее рассмотрим отдельно два случая.

1) $r' > r''$. В этом случае, обозначив $t = l + n$, учитывая условие нормировки $vu = 1$ и равенства $u^1 = u', u^2 = 0$, и пользуясь очевидным неравенством $u'vx^{(n)} \geq \min(u'v) \max x^{(n)}$ можно переписать неравенство $|A_t^* x - uvx^{(n)}| < \varepsilon \max x^{(n)}$ в виде:

$$\left|(A_t^* x)(1:k_1) - u'vx^{(n)}\right| \leq \varepsilon_1 u'vx^{(n)}, \ (A_t^* x)(k_1+1:k) \leq \varepsilon \max x^{(n)},$$

где $\varepsilon_1 = \varepsilon / \min(u'v)$. Домножая это же неравенство на v справа, получаем, что

$$|vA_t^* x - vx^{(n)}| < k\varepsilon \max(v) \max(x^{(n)}) \leq \varepsilon_2 vx^{(n)},$$

где $\varepsilon_2 = k\varepsilon \max(v) / \min(v)$.

Поэтому

$$\left|\frac{(A_t^* x)(1:k_1)}{vA_t^* x} - \frac{u'v \cdot x^{(n)}}{vx^{(n)}}\right| < \frac{u'vx^{(n)}}{vx^{(n)}} \cdot \frac{\varepsilon_1}{1-\varepsilon_2},$$
$$\frac{(A_t^* x)(k_1+1:k)}{vA_t^* x} < \frac{\varepsilon}{(1-\varepsilon_2)\min(v)},$$

(ввиду малости ε считаем, что $\varepsilon_1 < 1$). Отсюда получаем, что

$$\frac{A_t^* x}{vA_t^* x} = u + O(\varepsilon),$$

при $t > n_0 + l$. Устремляя ε к нулю, получаем требуемое.

2) $r'' > r'$. Тогда неравенство $|A_t^* x - uvx^{(n)}| < \varepsilon \max x^{(n)}$, подставляя $v^1 = 0, v^2 = v''$, можно записать в виде:

$$|A_t^* x - uv''x''^{(n)}| \leq \varepsilon_1 uv''x''^{(n)} + \varepsilon \max x'^{(n)},$$

$$|vA_t^* x - v''x''^{(n)}| \leq \varepsilon_2 v''x''^{(n)},$$

где

$$\varepsilon_1 = \varepsilon / \min(uv''), \ \varepsilon_2 = k\varepsilon \max(v'') / \min(v'')$$

Для того, чтобы доказать, что

$$\frac{A_t^* x}{vA_t^* x} = u + O(\varepsilon),$$

достаточно показать, что существует константа C, не зависящая от n, такая, что $\max x'^{(n)} < C \cdot \min x''^{(n)}$. Обозначим $A(n,j) = A_n A_{n-1} \dots A_{j+1}$.

Тогда можем записать, что

$$x''^{(n)} = (A_n^* x)(k_1+1:k) = A^{(3)}(n,0)x'',$$

$$x'^{(n)} = (A_n^* x)(1:k_1) = A^{(1)}(n,0)x' + \sum_{i=1}^{n-1} A^{(1)}(n,i) A_i^{(2)} A^{(3)}(i-1,0)x''.$$

Существует константа C_1 такая, что для любого i справедливо неравенство

$$A^{(1)}(i,0) < (A^{(1)})^i < C_1 \cdot (r')^i. \tag{III.2.9}$$

Пусть $r' < r'_1 < r''$. Возьмем такое n_0, что $A_n^{(3)} > r'_1 A^{(3)}/r''$ для $n > n_0$, где $r' < r'_1 < 1$. Это возможно, поскольку $A_n^{(3)} \to A^{(3)}$ при $n \to \infty$ и $A^{(3)} > 0$. Тогда при $i > n_0$ для некоторого положительного C_2 ввиду соотношения $(A^{(3)})^n/(r'')^n \to u''v''$ будет верно неравенство

$$A^{(3)}(n,i) > C_2 \cdot (r'_1)^{n-i}. \tag{III.2.10}$$

Поскольку n_0 - константа, а $A^{(3)}(i_1, i_2)$ положительны и ограничены при $i_1, i_2 < n_0$, то это неравенство справедливо для любого i для некоторой большей константы C_2.

Пусть элементы матриц $A_i^{(1)}, A_i^{(2)}, A_i^{(3)}$ ограничены сверху и снизу положительными константами M_1, M_2 соответственно.

Обозначим $C_3 = \max x / \min x''$. Поскольку

$$\min\Big(A^{(3)}(n,0)x''\Big) \geq k_2 \min\big(A^{(3)}(n,0)\big) \min(x'') > k_2 C_2 (r'_1)^n \min(x''),$$

то справедливо неравенство

$$\begin{aligned} A^{(1)}(n,0)x' < C_1 \cdot (r')^n k_1 \max(x') \leq C_1 \cdot (r'_1)^n C_3 k_1 \min(x'') \\ < \frac{C_1 C_3 k_1}{k_2 C_2} \cdot \min\big(A^{(3)}(n,0)x''\big). \end{aligned} \tag{III.2.11}$$

Кроме того, из неравенств (III.2.9), (III.2.10) и оценки $A_i^{(2)} x < k_2 M_2 \max x$ для любого вектора $x > 0$ следуют неравенства

$$\sum_{i=1}^{n-1} A^{(1)}(n,i) A_i^{(2)} A^{(3)}(i-1,0)x'' \leq k_1 \sum_{i=1}^{n-1} \max\Big(A^{(1)}(n,i)A_i^{(2)}\Big) \max(A^{(3)}(i-1,0)x'')$$

$$C_1 k_1 k_2 M_2 \sum_{i=1}^{n-1} \cdot (r')^i \max\Big(A^{(3)}(i-1,0)x''\Big)$$

$$\leq \frac{C_1 k_1 k_2 M_2}{C_2} \cdot \sum_{i=1}^{n-1} (r'/r'_1)^i \cdot \min\Big(A^{(3)}(n,0)x''\Big). \tag{III.2.12}$$

Здесь во второй оценке применено неравенство

$$\min(A^{(3)}(n,0)x'') \geq \min(A^{(3)}(n,i)) \max\Big(A^{(3)}(i,0)x''\Big) \geq C_2 (r'_1)^{n-i} \max\Big(A^{(3)}(i,0)x''\Big).$$

во

Складывая неравенства (III.2.11) и (III.2.12), получаем $\max x'_n < C \min x''_n$ для некоторой константы C. Очевидно, что C зависит только от матрицы A, скорости приближения $A_n \to A$, и константы C_3. Последнее нужно понимать в том смысле, что если мы нашли такую константу C для некоторой последовательности $A_n \to A$, то она будет действовать и для любой последовательности $A_n^* \geq A_n$.

Доказательство в случае $r' = r''$ проводится аналогично случаю $r' > r''$, с использованием утверждения (2) леммы III.2.2. Соотношение $\max\big((A_t^* x)(1 : k_1)\big) / \min\big((A_t^* x)(k_1 + 1 : k)\big) \to 0$ также следует из утверждения (2) леммы, а именно, из оценок

$$(A^t x)(1 : k_1) \sim u'v'x' + tu'v''x'', \ (A^t x)(k_1 + 1 : k) \sim u''v''x''$$

при $t \to \infty$.

Заметим, что сходимость в соотношениях (III.2.4), (III.2.5) равномерна по x (в случае $r' < r''$ - при условии $\max x \le C_3 \min x''$ для заданной константы C_3), и для любой последовательности $A_n^* \ge A_n$, где последовательности A_n^*, A_n удовлетворяют условиям леммы.

Пусть $F(t,s) = (F_1(t,s), \dots, F_k(t,s))$ - (векторная) производящая функция, которая определяется как t-я итерация производящей функции $F(s)$ соотношениями:

$$F(0,s) = s, F(1,s) = F(s),$$

$$F(t+1,s) = F(F(t,s)). \tag{III.2.13}$$

Очевидно, что $F(t,\bar{1}) = \bar{1}$ для всех t. Из [16] известно, что $Q(t) = \bar{1} - F(t,\bar{0})$, где $Q(t) = (Q_1(t), \dots, Q_k(t))$. Обозначим $R(t,s) = \bar{1} - F(t,s)$, в частности, $R(t,\bar{0}) = Q(t)$.

Лемма III.2.4 *Для стохастической КС-грамматики с матрицей первых моментов A вида (I.2.1) при $r \le 1$ верно равенство*

$$\bar{1} - F(s) = [A - E(s)](\bar{1} - s), \tag{III.2.14}$$

где $0 \le E(s) \le A$, причем элементы матрицы $E(s)$ при $0 \le s \le 1$ удовлетворяют условиям

$$E_{ij}(s) = \sum_l \delta_{jl}^i(s)(1 - s_l), \text{ где } 0 \le \delta_{jl}^i(s) \le b_{jl}^i \tag{III.2.15}$$

и $E(s)$ имеет блочный вид (I.2.1) при любом $0 \le s \le 1$.

Доказательство. Используя разложение производящей функции $F_i(s)$ в ряд Тейлора в окрестности 1, мы можем записать:

$$1 - F_i(s) = \sum_j \frac{\partial F_i(s)}{\partial s_j}|_{s=\theta^i}(1 - s_j) = \sum_j E_{ij}(s)(1 - s_j),$$

где $\theta^i = (\theta_1^i, \dots \theta_k^i)$, причем $\theta^i = \alpha^i \cdot \bar{1} + (1 - \alpha^i) \cdot s$ $(0 \le \alpha^i \le 1)$. Поскольку производящие функции $F_i(s)$ - многочлены с положительными коэффициентами, все их производные монотонно возрастают по всем s_j, и, следовательно, $0 \le \frac{\partial F_i(s)}{\partial s_j} \le \frac{\partial F_i(s)}{\partial s_j}|_{s=\theta^i} \le a_j^i$. Раскладывая $\frac{\partial F_i(s)}{\partial s_j}$ аналогичным образом в ряд Тейлора, имеем: $\frac{\partial F_i(s)}{\partial s_j} = a_j^i - \sum_l \delta_{jl}^i(s)(1 - s_l)$, где $0 \le \delta_{jl}^i(s) \le b_{jl}^i = \frac{\partial^2 F_i(s)}{\partial s_j \partial s_l}|_{s=1}$. Отсюда и следуют равенства (III.2.14) и (III.2.15). Утверждение о том, что матрица $E(s)$ имеет блочный вид (I.2.1), следует из того, что $a_j^i = b_{jl}^i = 0$ при $i > k_1$ и $j \le k_1$.

Подставляя в качестве s вектор $F(t,s)$ в соотношение (III.2.14) и используя равенство (III.2.13), получаем:

$$\bar{1} - F(t+1,s) = (A - E(F(t,s)))\,(\bar{1} - F(t,s)). \tag{III.2.16}$$

Обозначим $E(F(t,s))$ за $E_t(s)$, а $E_t(0)$ за E_t и применим формулу (III.2.16) рекурсивно. Тогда:

$$R(t,s) = \bar{1} - F(t,s) = \prod_{i=1}^{t-1}(A - E_i(s))(\bar{1} - s), \tag{III.2.17}$$

откуда при $s = \bar{0}$ имеем

$$Q(t) = \prod_{i=1}^{t-1}(A - E_i) \cdot \bar{1}. \qquad \text{(III.2.18)}$$

Кроме того, из формулы (III.2.17) следует, что для любого $0 \le s \le 1$

$$R(t,s) = \bar{1} - F(t,s) \le A^t \cdot (\bar{1} - s). \qquad \text{(III.2.19)}$$

Так как $A^t = O(tr^t)$, то $F(t,s) \to 1$ при $t \to \infty$, и, следовательно, $\lim\limits_{t\to\infty} E_t(s) = 0$ по лемме III.2.4 (поэлементно).

Лемма III.2.5 *Пусть матрица первых моментов грамматики имеет вид (I.2.1) и ее перронов корень $r < 1$. Тогда при любом $0 \le s \le 1$ справедливы следующие утверждения:*
1) при $r' \ne r''$ верна оценка $E_t(s) = O(r^t)$;
2) при $r' = r''$ верна оценка $E_t(s) = O(tr^t)$.

Доказательство. Сначала докажем утверждение (1). Из неравенства (III.2.19) и леммы III.2.2 следует, что при $r' \ne r''$ верна оценка $R(t,s) = O(r^t)$. По лемме III.2.4 справедлива оценка

$$E_t(s)_{ij} = E(F(t,s))_{ij} \le \sum_l \delta^i_{jl}(F(t,s))(1 - F_l(t,s)) = O(r^t).$$

Отсюда вытекает первое утверждение леммы. Второе доказывается таким же образом с использованием оценки $R(t,s) = O(tr^t)$ при $r' = r''$, которая также следует из формулы (III.2.19).

Доказательство теоремы III.2.1.
Рассмотрим сначала случай $r' \ne r''$. Ввиду равенства (III.2.18) и леммы III.2.3 верно соотношение $\lim\limits_{t\to\infty} \frac{Q(t)}{vQ(t)} = u$, из которого следует асимптотическое равенство $Q_i(t) = u_i vQ(t)(1 + o(1))$. Следовательно, достаточно показать, что существует конечный $\lim\limits_{t\to\infty} vQ(t)/r^t > 0$. Поскольку $Q(t+1) = (A - E_t)Q(t)$, то $vQ(t+1) = r \cdot vQ(t) - vE_tQ(t)$, так что $vQ(t+1) = r\left(1 - \frac{vE_tQ(t)}{rvQ(t)}\right) \cdot vQ(t)$. Ввиду доказанного в лемме III.2.4 неравенства $0 \le E_t(s) \le A$, доказываемое утверждение следует из сходимости бесконечного произведения $\prod\limits_{t=0}^{\infty}\left(1 - \frac{vE_tQ(t)}{rvQ(t)}\right)$, что по известному критерию [18] сходимости бесконечного произведения следует из сходимости ряда $\sum\limits_{t=0}^{\infty} \frac{vE_tQ(t)}{rvQ(t)}$. Сходимость этого ряда следует из оценки $E_t(s) = O(r^t)$, доказанной в лемме III.2.5.

Теперь рассмотрим случай $r = r' = r''$. Обозначим $Q(t)(1 : k_1) = x'_t$, $Q(t)(k_1 + 1 : k) = x''_t$. Используя известную из [16] формулу для вероятностей продолжения неразложимого ветвящегося процесса, соответствующего классу K_2, получаем, что для $i > k_1$ при $t \to \infty$ выполняется асимптотическое равенство

$$Q_i(t) = c_0 u''_{i-k_1}(r'')^t(1 + o(1)), \qquad \text{(III.2.20)}$$

где $c_0 > 0$. Поэтому $x''_t = c_0 u'' r^t(1+o(1))$.

Пусть
$$E_t = \begin{pmatrix} E_t^{(1)} & E_t^{(2)} \\ 0 & E_t^{(3)} \end{pmatrix}.$$

Домножая равенство $x'_{t+1} = (A - E_t^{(1)})x'_t + (A^{(2)} - E_t^{(2)})x''_t$ слева на v' и подставляя выражение для x_t^2, получаем, что $v'x'_{t+1} = rv'x'_t - v'E_t^{(1)}x'_t + v'(A^{(2)} - E_t^{(2)})u'' \cdot c_0 r^t(1+o(1))$. Обозначим $x_t = v'x'_t/r^t$. Тогда предыдущее равенство переписывается таким образом:

$$x_{t+1} = x_t + v'(A^{(2)} - E_t^{(2)})u'' \cdot r^{-1}c_0(1+o(1)) - v'E_t^{(1)}x'_t/r^{t+1}. \qquad \text{(III.2.21)}$$

Поскольку из равенства (III.2.21) вытекает, что $x_{t+1} \le x_t + r^{-1}v'A^{(2)}u'' \cdot c_0(1+o(1))$, то существует константа $C > 0$, такая что $x_t < Ct$. Отсюда, используя оценку $E_t = O(tr^t)$ леммы III.2.5, находим, что $v'E_t^{(1)}x'_t/r^{t+1} = O(t^2r^t) \to 0$ при $t \to \infty$. Следовательно, $x_t = c_0 v'A^{(2)}u''t/r + o(t)$. Используя вторую часть утверждения леммы III.2.3, получаем требуемое.

Следствие III.2.2 *Вероятность деревьев вывода высоты t при $t \to \infty$ имеет следующий вид:*
1) при $r' \ne r''$
$$P_i(t) = c_0 u_i \cdot (1-r)r^{t-1} + o(r^t);$$
2) при $r' = r''$
$$P_i(t) = bc_0 u'_i \cdot (1-r)tr^{t-1} + o(tr^t),\ i = 1, \ldots, k_1,$$
$$P_i(t) = c_0 u''_{i-k_1} \cdot (1-r)r^{t-1} + o(r^t),\ i = k_1+1, \ldots, k,$$
где b введено формулой (III.2.3).

III.3 Закономерности в деревьях вывода слов. Случай простого перронова корня

В этом параграфе мы рассмотрим случай $r' \ne r''$, $r = \max(r', r'') < 1$. Введем случайную величину $q_{ij}(t, \tau)$ - число применений правила r_{ij} на ярусе τ в деревьях вывода высоты t. Обозначим ее математическое ожидание через $M(q_{ij}(t,\tau))$. Поскольку вероятности продолжения $Q_i(t)$ и первые моменты в случае $r' \ne r''$ для рассматриваемой разложимой грамматики имеют тот же вид, что и для неразложимой, все доказанные в [9] для неразложимого случая результаты относительно математического ожидания числа применений правила грамматики на фиксированном ярусе в деревьях вывода заданной высоты и математического ожидания числа применений правила грамматики на одном ярусе дерева вывода остаются в силе. Поэтому справедливы следующие результаты.

Теорема III.3.1 *Пусть матрица первых моментов КС-грамматики имеет вид (I.2.1), ее перронов корень $r < 1$ и $r' \neq r''$. Тогда при $\tau \to \infty,\ t-\tau \to \infty$ выполняется соотношение*

$$M(q_{ij}(t,\tau)) \to p_{ij} \cdot \left(\frac{v_i}{r} \sum_{l=1}^{k} u_l s_{ij}^l + \sum_{l=1}^{k} u_l g_{il} \right),$$

где u, v - соответственно правый и левый собственные вектора матрицы первых моментов, соответствующие перронову корню, s_{ij}^l - число нетерминалов A_l в правой части правила r_{ij}, а g_{ij} - константы, определяемые равенствами

$$b_{jn}^i(t) = u_i g_{jn} r^t + O(r_1^t),\ r_1 < r,$$

в которых $b_{jn}^i(t)$ - вторые моменты грамматики.

Обозначим за $M(x_i(t,\tau))$ математическое ожидание числа нетерминалов A_i на ярусе τ в деревьях вывода высоты t.

Следствие III.3.1 *Для КС-грамматики, матрица первых моментов которой имеет вид (I.2.1), ее перронов корень $r < 1$ и $r' \neq r''$, при $\tau \to \infty,\ t - \tau \to \infty$ выполняется соотношение:*

$$M(x_i(t,\tau)) \to u_i v_i + \sum_{l=1}^{k} u_l g_{il}.$$

Обозначим через $M(\omega(t,\tau))$ математическое ожидание числа нетерминалов на ярусе τ в деревьях вывода высоты t.

Следствие III.3.2 *Пусть матрица первых моментов КС-грамматики имеет вид (I.2.1), ее перронов корень $r < 1$ и $r' \neq r''$. Тогда при $\tau \to \infty,\ t-\tau \to \infty$ справедливо соотношение*

$$M(\omega(t,\tau)) \to 1 + \sum_{i=1}^{k} \sum_{m=1}^{k} u_m g_{im}.$$

Далее, пусть $S_{ij}(t) = q_{ij}(t,1) + q_{ij}(t,2) + \ldots + q_{ij}(t,t)$, то есть $S_{ij}(t)$ - число применений правила r_{ij} во всем дереве вывода. Асимптотика этой величины описывается следующей теоремой:

Теорема III.3.2 *Пусть матрица первых моментов КС-грамматики имеет вид (I.2.1), ее перронов корень $r < 1$ и $r' \neq r''$. Тогда $M(S_{ij}(t)/t) \to p_{ij} \cdot \left(\frac{v_i}{r} \sum\limits_{l=1}^{k} u_l s_{ij}^l + \sum\limits_{l=1}^{k} u_l g_{il} \right)$ при $t \to \infty$.*

III.4 Закономерности в деревьях вывода слов. Случай кратного перронова корня

В этом параграфе мы рассмотрим случай $r' = r'' < 1$. Введем следующие обозначения. Пусть $X = (x_1, \dots, x_k)$ - вектор с неотрицательными целыми компонентами. Через $D_X(\tau)$ обозначим множество деревьев вывода, имеющих x_i нетерминалов A_i на ярусе τ (где $i = 1, \dots, k$), а через $D_X^{\leq t}(\tau)$ - множество всех деревьев вывода из $D_X(\tau)$, имеющих высоту не более t. Подмножество деревьев из $D_X^{\leq t}(\tau)$ высоты ровно t обозначим $D_X^t(\tau)$. Через $D_X^{\leq n}$ обозначим множество наборов деревьев вывода

$$d_X^{\leq n} = (d_{11}, \dots, d_{1x_1}, \dots, d_{k1}, \dots, d_{kx_k}),$$

где корни деревьев $d_{m1}, \dots, d_{mx_m}$ помечены нетерминалом A_m, $m = 1, \dots, k$, причем каждое дерево в наборе имеет высоту не больше n. Через $p(d_X^{\leq n})$ обозначим произведение вероятностей деревьев набора d. Через $\mathcal{D}_X^n$ обозначим множество наборов из $D_X^{\leq n}$, которые содержат хотя бы одно дерево высоты n. Наборы из D_X^n будем обозначать d_X^n, а произведение вероятностей деревьев, образующих d_X^n, за $p(d_X^n)$. Обозначим $Q_X(n) = \sum p(d_X^{\leq n})$, где суммирование производится по всем наборам из $D_X^{\leq n}$. Через $R_X(n)$ обозначим $\sum p(d_X^n)$, где суммирование производится по всем наборам из $\mathcal{D}_X^n$. Ввиду согласованности грамматики, справедливо равенство $P(D_X^t(\tau)) = P(D_X(\tau)) \cdot R_X(t - \tau)$, поскольку для такой грамматики вероятность $P(D_X(\tau))$ равна суммарной вероятности всех деревьев высоты τ, имеющих на ярусе τ вектор X нетерминалов. Далее $P(D_X^t(\tau))$ будем обозначать для краткости через $P_X(\tau)$. Произведем оценку величин $R_X(n)$ и $Q_X(n)$. Для оценки $Q_X(n)$ нам понадобится следующая лемма из [16]:

Лемма III.4.1 *Пусть s, d - натуральные, $n_1, \dots, n_s$, $m = (m_1, \dots, m_s)$ - целые неотрицательные числа, $y = (y_1, \dots, y_s)$. Тогда*

$$(1 - y_1)^{n_1} \dots (1 - y_s)^{n_s} = \sum_{\bar{m} < d, m \geq 0} C_{n_1}^{m_1} \dots C_{n_s}^{m_s} (-1)^{\bar{m}} y^m$$

$$+ R_d(n_1, \dots, n_s; y),$$

где остаточный член представим в виде

$$R_d(n_1, \dots, n_s, y) = \sum_{\bar{m} = d, m \geq 0} (-1)^d \varepsilon_m(n_1, \dots, n_s; y) \cdot y^m,$$

причем

$$0 \leq \varepsilon_m(n_1, \dots, n_s; y') \leq \varepsilon_m(n_1, \dots, n_s; y) \leq C_{n_1}^{m_1} \dots C_{n_s}^{m_s}$$

при $0 \leq y_i \leq y'_i \leq 1$ $(i = 1, \dots, s)$.

Используя эту лемму при $d = 2$ и полученные в теореме III.2.1 выражения для $Q_i(t), i = 1, \dots, k$, имеем:

$$Q_X(n) = \prod_{i=1}^k (1 - Q_i(n))^{x_i} = 1 - \sum_{i=1}^k x_i Q_i(n) + O\left(\sum_{i,j} x_i x_j Q_i(n) Q_j(n)\right)$$

$$= 1 - c_0 \cdot \left(\sum_{i=1}^{k_1} x_i u'_i nb + \sum_{i=k_1+1}^{k} x_i u''_i \right) \cdot r^n (1 + o(1)) + O(r^{2n}) Y_X(n), \qquad \text{(III.4.1)}$$

где

$$Y_X(n) = \sum_{i,j=1}^{k} x_i x_j Q_i(n) Q_j(n) / r^{2n} \le O\left(\sum_{i,j} x_i x_j n^2 \right).$$

Иногда нам будет достаточно более грубой оценки, которая является очевидным следствием равенства (III.4.1):

$$Q_X(n) = 1 - O(r^n) Z_X(n), \qquad \text{(III.4.2)}$$

где

$$Z_X(n) = \sum_{i=1}^{k} x_i Q_i(n) / r^n \le O\left(\sum_{i} x_i n \right).$$

Чтобы вывести формулу для $R_X(n)$, используем доказанную в [9] оценку:

$$R_X(n) = \sum_{i=1}^{k} x_i \cdot (Q_i(n-1) - Q_i(n)) + O(r^{2n} Y_X(n)).$$

Пользуясь этим соотношением, мы можем записать, что

$$R_X(n) = c_0 (1-r) r^{n-1} \left(\sum_{i=1}^{k_1} x_i u'_i bn + \sum_{i=k_1+1}^{k} x_i u''_{i-k_1} \right) (1 + o(1)) + O(r^{2n}) Y_X(n). \quad \text{(III.4.3)}$$

III.5 Оценки моментов второго и более высоких порядков для кратного перронова корня

Для нахождения математического ожидания числа применений правила на ярусе дерева вывода нам понадобится вид первых и вторых моментов грамматики. Вид первых моментов $a^i_j(t)$ получен в следствии III.2.1. Вид вторых моментов описывается следующей леммой.

Лемма III.5.1 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) при $r = r' = r'' < 1,\ t \to \infty$, вторые моменты $b^i_{js}(t)$ имеют следующий вид:*
$b^i_{js}(t) = u'_i g_{js} r^t (1 + o(1))$, *при $i \le k_1$, если $j \le k_1$ либо $s \le k_1$;*
$b^i_{js}(t) = b u'_i g_{js} \cdot t r^t (1 + o(1))$, *при $i \le k_1$ и $j, s > k_1$;*
$b^i_{js}(t) = u''_{i-k_1} g_{js} r^t (1 + o(1))$ *при $i, j, s > k_1$;*
$b^i_{js}(t) = 0$ *при $i > k_1$, если $j \le k_1$ или $s \le k_1$.*
Здесь g_{js} - некоторые константы, причем $g_{js} = g_{sj},\ j, s = 1, \ldots, k$.

Доказательство. Для вторых моментов в [16] доказано соотношение

$$b_{jn}^i(t) = \sum_{\tau=1}^{t} \sum_{l,m,s} a_l^i(t-\tau) \cdot b_{ms}^l a_j^m(\tau-1) a_n^s(\tau-1). \tag{III.5.1}$$

Поскольку $A(t) = a_j^i(t) = A^t$, а в [8] доказано, что в случае, если перронов корень r матрицы A имеет кратность два, верно равенство :

$$A^t = (A_0 t + A_1) r^t + \sum_{q=1}^{h} \phi_{jq}^i(t) r_q^t,$$

где $r_q, q = 1, \ldots, h$ - собственные корни, отличные от r, $|r_q| < r$, и $\phi_{jq}^i(t)$ - полином, степень которого меньше, чем кратность корня r_q. В частности, для собственного корня r этот полином выписан здесь явно как $A_0 t + A_1$, где A_0, A_1 - матрицы размера $k \times k$. При этом в лемме III.2.2 мы доказали, что:

$$\begin{array}{l} A_{0j}^i = b u_i' v_{j-k_1}''/r \text{ при } i \le k_1 \text{ и } j > k_1, \\ A_{0j}^i = 0 \text{ в остальных случаях,} \end{array}$$

$$\begin{array}{l} A_{1j}^i = u_i' v_j' \text{ при } i, j \le k_1, \\ A_{1j}^i = u_{i-k_1}'' v_{j-k_1}'' \text{ при } i, j > k_1, \\ A_{1j}^i = 0 \text{ при } i > k_1 \text{ и } j \le k_1. \end{array}$$

Мы ограничимся рассмотрением случая (3): $j, n > k_1$, $i \le k_1$, остальные случаи рассматриваются полностью аналогично. Преобразуя правую часть равенства (III.5.1), получаем

$$b_{jn}^i(t) = \sum_{\tau=1}^{t} \sum_{l \le k_1, 0 < m,s \le k} a_l^i(t-\tau) \cdot b_{ms}^l a_j^m(\tau-1) a_n^s(\tau-1)$$

$$+ \sum_{\tau=1}^{t} \sum_{l,m,s > k_1} a_l^i(t-\tau) \cdot b_{ms}^l a_j^m(\tau-1) a_n^s(\tau-1).$$

остальные слагаемые равны нулю, поскольку $a_j^i(t) = 0$ при $i > k_1, j \le k_1$, и $b_{js}^i(t) = 0$ при $i > k_1$ и $j \le k_1$ или $s \le k_1$. Первое слагаемое оценивается как

$$\sum_{\tau=1}^{t} \sum_{l \le k_1, m,s} a_l^i(t-\tau) \cdot b_{ms}^l a_j^m(\tau-1) a_n^s(\tau-1) = O\left(\sum_{\tau=1}^{t} A_{1j}^i \tau^2 \cdot r^{t-\tau} r^{2\tau} \right) = O(r^t).$$

Здесь использовался тот факт, что для $0 < r < 1, n \ge 0,\ t \to \infty$

$$\sum_{i=n}^{t} C_i^n r^{i-n} = \frac{1 + O(t^n r^t)}{(1-r)^{n+1}}. \tag{III.5.2}$$

Это соотношение доказывается путем n-кратного дифференцирования по r равенства $\sum\limits_{i=0}^{t} r^i = (1-r^t)/(1-r)$.

Второе слагаемое можно переписать как

$$\sum_{\tau=1}^{t}\sum_{l,m,s>k_1} a_l^i(t-\tau)\cdot b_{ms}^l a_j^m(\tau-1)a_n^s(\tau-1) = \sum_{\tau=1}^{t}\sum_{l,m,s>k_1}\left[bu_i' v_{j-k_1}'' \cdot tr^{t-\tau-1} + O(r^t)\right]$$

$$\times b_{ms}^l \times \left[A_{1j}^m r^{\tau-1} + \sum_{q=1}^{h}\phi_{jq}^m(\tau-1)r_q^{\tau-1}\right] \times \left[A_{1n}^s r^{\tau-1} + \sum_{q=1}^{h}\phi_{nq}^s(\tau-1)r_q^{\tau-1}\right]$$

$$= tr^{t-1}bu_i' \cdot \sum_{\tau=1}^{t}\sum_{l,m,s>k_1} v_{j-k_1}'' b_{ms}^l r^{-1} \times \left[A_{j1}^m + \sum_{q=1}^{h}\phi_{jq}^m(\tau-1)r_q^{\tau-1}\right]$$

$$\times \left[A_{1n}^s r^{\tau-1} + \sum_{q=1}^{h}\phi_{nq}^s(\tau-1)r_q^{\tau-1}\right] + O(r^t).$$

Ряд

$$\sum_{\tau=1}^{t}\sum_{l,m,s>k_1} v_{j-k_1}'' b_{ms}^l r^{-1} \times \left[A_{1j}^m + \sum_{q=1}^{h}\phi_{jq}^m(\tau-1)r_q^{\tau-1}\right]$$

$$\times \left[A_{1n}^s r^{\tau-1} + \sum_{q=1}^{h}\phi_{nq}^s(\tau-1)r_q^{\tau-1}\right] \tag{III.5.3}$$

сходится. Действительно,

$$v_{j-k_1}'' b_{ms}^l r^{-1} \times \left[A_{1j}^m + \sum_{q=1}^{h}\phi_{jq}^m(\tau-1)r_q^{\tau-1}\right]$$

мажорируется константой для всех $\tau > 1$, так как $\phi_{jq}^m(\tau-1)$ - полиномы, а $r < 1$. Ряд

$$\sum_{\tau=1}^{t}\left[A_{1n}^s r^{\tau-1} + \sum_{q=1}^{h}\phi_{nq}^s(\tau-1)r_q^{\tau-1}\right]$$

сходится ввиду соотношений (III.5.2). Обозначив сумму ряда (III.5.3) через g_{jn}, получаем доказательство рассмотренного утверждения леммы.

Для случая $k_1 = k_2 = 1$ величины g_{ij} могут быть выражены явно с помощью формул (III.5.1). Введем следующие обозначения:

$$A(\tau) = (a_j^i(\tau)),\ B_{jn}^l(\tau) = b_{ms}^l a_j^m(\tau)a_n^s(\tau),\ B^l(\tau) = (B_{ij}^l(\tau)).$$

Обозначив $k_\tau = b\cdot(\tau-1) = a_2^1\cdot(\tau-1)/r$ и, учитывая, что $a_1^1(t) = a_2^2(t) = r^t, a_2^1(t) = k_t r^t, a_1^2(t) = 0$, мы можем записать:

$$B^l(\tau-1) = r^{2(\tau-1)}\cdot\begin{pmatrix} b_{11}^l & b_{11}^l k_\tau + b_{12}^l \\ b_{11}^l k_\tau + b_{12}^l & b_{22}^l + 2b_{12}^l k_\tau + k_\tau^2 b_{11}^l \end{pmatrix}.$$

Учитывая, что $b^l_{ij}(t) = b^l_{ji}(t)$, $b^2_{ij}(t) = 0$ при $i \neq 2$ или $j \neq 2$, получаем, что $B^2_{jn}(\tau-1) = 0$ для $i \neq 2$ или $j \neq 2$, а

$$B^2_{22}(\tau-1) = r^{2(\tau-1)} b^2_{22}.$$

Используя формулу (III.5.1) и равенства (III.5.2), находим :

$$b^1_{11}(t) = \sum_{\tau=1}^{t} r^{t-\tau} b^1_{11} = b^1_{11} \frac{r^{t-1}}{1-r}(1 + O(r^t));$$

$$\begin{aligned} b^1_{12}(t) &= \sum_{\tau=1}^{t} r^{t-\tau} \cdot \left(b^1_{12} + b^1_{11} b \cdot (\tau-1)\right) r^{2(\tau-1)} \\ &= \frac{r^{t-1}}{1-r} \cdot \left(b^1_{12} + \frac{b^1_{11} a^1_2}{1-r}\right)(1 + O(tr^t)); \end{aligned}$$

$$b^2_{22}(t) = \sum_{\tau=1}^{t} a^2_2(t-\tau) B^2_{22}(\tau-1) = \sum_{\tau=1}^{t} r^{t-\tau} \cdot r^{2(\tau-1)} b^2_{22} = \frac{b^2_{22}}{1-r} r^{t-1}(1 + O(r^t)).$$

Поэтому для $k_1 = k_2 = 1$, $a^1_1 = a^2_2 = r$, верны равенства:

$$\begin{aligned} g_{11} &= \frac{b^1_{11}}{r(1-r)}, \; g_{22} = \frac{b^2_{22}}{r(1-r)}, \\ g_{12} &= \frac{1}{r(1-r)} \cdot \left(b^1_{12} + \frac{b^1_{11} a^1_2}{1-r}\right). \end{aligned} \tag{III.5.4}$$

Для моментов более высокого порядка справедлива

Лемма III.5.2 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) при $r = r' = r'' < 1$ и $t \to \infty$ имеет место оценка $m^i_\alpha(t) = O(r^t)$, кроме случая, когда $i \leq k_1$ и $\alpha > k_1$, для которого верна оценка $m^i_\alpha(t) = O(tr^t)$.*

Лемма доказывается индукцией по $\bar{\alpha}$. Для $\bar{\alpha} = 1, 2$ утверждение уже доказано. Допустим, что оно верно для всех $\bar{\alpha} < n$. Докажем его для $\bar{\alpha} = n$. Очевидно, что $m^i_\alpha(t) = 0$ при $i > k_1$, если не выполнено $\alpha_s > k_1$ для всех $s = 1, \ldots, k$. В последнем случае $m^i_\alpha(t)$ имеет вид $O(r^t)$ (для доказательства достаточно рассмотреть неразложимую грамматику, порожденную нетерминалами из класса K_2), поскольку ее моменты $m'^i_\alpha = m^{i-k_1}_{\alpha-k_1}$.

Рассмотрим случай $i \leq k_1$. В [16] доказано соотношение

$$m^i_\alpha(t) = \sum_{\tau=1}^{t} \sum_{s} a^i_s(t-\tau) z^s_\alpha(\tau-1),$$

причем

$$z^s_\alpha(t) = m^s_\alpha(t+1) - \sum_n a^s_n m^n_\alpha(t).$$

Можно записать, что

$$m^i_\alpha(t) = \sum_{\tau=1}^{t} \sum_{0<s\leq k_1} a^i_s(t-\tau) z^s_\alpha(\tau-1) + \sum_{\tau=1}^{t} \sum_{k_1<s<k} a^i_s(t-\tau) z^s_\alpha(\tau-1).$$

Поэтому, если хотя бы один индекс $\alpha_i \le k_1$, то при $s > k_1$

$$z^s_\alpha(t) = m^s_\alpha(t+1) - \sum_n a^s_n m^n_\alpha(t) = 0.$$

При этом $z^s_\alpha(\tau-1)$ является суммой произведений моментов более низких порядков: $m^i_\beta m^{i_1}_{\gamma_1}(t)\cdot\ldots\cdot m^{i_{\bar\beta}}_{\gamma_{\bar\beta}}(t)$, где $1<\bar\beta\le\bar\alpha$, $\sum_m \gamma_m=\alpha$, $\gamma_m\ne\alpha, 1\le i_m\le k$. Используя оценки для моментов более низких порядков, для первой суммы получаем оценку $O(r^t)$. Для второй суммы члены, которые могли бы дать порядок $O(tr^t)$ - это $a^i_s z^s_\alpha(t)$ при $i\le k_1$ и $s>k_1$, но они равны нулю, если $\alpha_i\le k_1$ хотя бы для одного i. При $\alpha>k_1$ эти члены имеют порядок $O(t\cdot r^t)$. Доказательство завершено.

Исследуем асимптотическое поведение $M(q_{ij}(t,\tau))$ при $\tau, t-\tau\to\infty$. Зафиксируем вектор $X=(x_1,\ldots,x_k)$ и найдем вклад деревьев вывода из $D^t_X(\tau)$ в $M(q_{ij}(t,\tau))$. Он равен $\Delta_X=\frac{1}{P(D^t)}x_i\cdot(P^I+P^{II})$, где

$$P^I = P_X(\tau)p_{ij}R_S(t-\tau-1)Q_{X'}(t-\tau)$$

- суммарная вероятность деревьев из $D^t_X(\tau)$, для которых последний ярус достигается через применение правила r_{ij} к первому по порядку нетерминалу A_i на ярусе τ, и

$$P^{II} = P_X(\tau)p_{ij}R_{X'}(t-\tau)Q_S(t-\tau-2)$$

- суммарная вероятность деревьев из $D^t_X(\tau)$, для которых последний ярус достигается благодаря применению других правил (т.е. поддерево с корнем, помеченным самым левым нетерминалом A_i на ярусе τ, имеет высоту меньше $t-\tau$). Здесь $X'=(x_1,\ldots,x_{i-1},x_i-1,x_{i+1},\ldots,x_k)$, и $S=(s^1_{ij},\ldots,s^k_{ij})$, где s^l_{ij} - число нетерминалов A_l в правой части правила r_{ij}.

Рассмотрим сначала случай $i\le k_1$. Найдем $S_1=R_S(t-\tau-1)Q_{X'}(t-\tau)$. Используя полученные выше оценки (III.4.2), (III.4.3) для $R_X(n)$ и $Q_X(n)$, получаем, что

$$S_1 = c_0\cdot(1-r)r^{t-\tau-2}\times\left(b\cdot(t-\tau-1)\sum_{l=1}^{k_1}u'_l s^l_{ij}+\sum_{l=k_1+1}^{k}u''_{l-k_1}s^l_{ij}\right)$$
$$\times\Big(1+O(r^{t-\tau}\cdot Z_X(t-\tau-1))\Big)(1+o(1)),$$

поскольку погрешность для $R_S(t-\tau-1)$ не зависит от X и может быть записана как $o(1)$. Если обозначить $\sum\limits_{l=1}^{k_1}u'_l s^l_{ij}=S'_{ij}$, $\sum\limits_{l=k_1+1}^{k}u''_{l-k_1}s^l_{ij}=S''_{ij}$, то выражение для S_1 перепишется в виде

$$S_1 = c_0\cdot(1-r)r^{t-\tau-2}\cdot\Big(b\cdot(t-\tau-1)S'_{ij}+S''_{ij}\Big)(1+o(1))+O\Big(r^{t-\tau}\cdot Z_X(t-\tau-1)\Big).$$

Аналогично оценим $S_2=R_{X'}(t-\tau)Q_S(t-\tau-2)$. Очевидно,

$$S_2 = c_0\cdot(1-r)r^{t-\tau-1}\times\Big(\big(b\cdot(t-\tau)\sum_{l=1}^{k_1}x'_l u'_l+\sum_{l=k_1+1}^{k}x'_l u''_{l-k_1}\big)$$

$$+O(r^{t-\tau}\cdot\ Y_X(t-\tau-2))\Big)(1+o(1)),$$

где $x'_l = x_l$ при $l \neq i$, $x'_i = x_i - 1$. Здесь $Q_S(t-\tau-2)$ оценено как $1+o(1)$ с использованием оценки (III.4.2). Подставляя формулы для S_1, S_2 в выражения для P^I, P^{II} и суммируя по X, а также учитывая вытекающие из следствия III.2.1 и лемм III.5.1, III.5.2 оценки

$$\sum_X P_X(\tau)x_i Z_X(t-\tau-1) = O((t-\tau)\cdot r^\tau),$$

$$\sum_X P_X(\tau)x_i Y_X(t-\tau-2) = O((t-\tau)^2\cdot r^\tau)$$

для $i \le k_1$, получаем, что при $\tau, t-\tau\to\infty$

$$M(q_{ij}(t,\tau)) = \frac{1}{P(D^t)}\sum_X x_i P_X(\tau)p_{ij}\cdot(S_1+S_2)$$

$$= \frac{p_{ij}c_0\cdot(1-r)r^{t-\tau-2}(1+o(1))}{c_0btu'_1(1-r)r^{t-1}} \times \Bigg[\sum_X P_X(\tau)x_i S'_{ij} b\cdot(t-\tau-1) + \sum_X P_X(\tau)x_i S''_{ij}$$

$$+\sum_X P_X(\tau)\cdot\sum_{l=1}^{k_1} u'_i x_i x'_l\cdot br\cdot(t-\tau) + r\cdot\sum_X P_X(\tau)\cdot\sum_{l=k_1+1}^{k} u''_{l-k_1}x_i x_l\Bigg].$$

Поскольку

$$\sum_X P_X(\tau)x_i = a_i^1(\tau) = O(r^\tau),\ \sum_X P_X(\tau)x_i x_l = b_{il}^1(\tau) = O(r^\tau),$$

то членами $S''_{ij}\cdot\sum\limits_X P_X(\tau)x_i$ и $r\cdot\sum\limits_{l=k_1+1}^{k} u''_{l-k_1}b_{il}^1(\tau)$ можно пренебречь при $t-\tau\to\infty$, и, следовательно,

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}}{u'_1 tr^{\tau+1}}\cdot\left(a_i^1(\tau)\cdot(t-\tau-1)S'_{ij} + br(t-\tau)\cdot\sum_{l=1}^{k_1} u'_l b_{il}^1(\tau)\right)(1+o(1)).$$

Преобразуя выражение для $M(q_{ij}(t,\tau))$ дальше и подставляя выражения для первых и вторых моментов, полученные в следствии III.2.1 и лемме III.5.1, имеем, что

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}}{t}\cdot\left(\frac{(t-\tau-1)v'_i S'_{ij}}{r} + \sum_{l=1}^{k_1} u'_l g_{il}\cdot(t-\tau)\right)(1+o(1))$$

$$= \frac{p_{ij}\cdot(t-\tau)}{t}(G'_i + v'_i S'_{ij}/r)(1+o(1)),$$

где $G'_i = \sum\limits_{l=1}^{k_1} u'_l g_{il}$.

Теперь рассмотрим случай $i > k_1$. Поскольку $s_{ij}^l = 0$ при $i > k_1$ и $l \le k_1$, то $S'_{ij} = 0$. Используя оценки

$$\sum_X P_X(\tau) x_i Z_X(t-\tau-1) = O(\tau(t-\tau)\cdot r^\tau),$$

$$\sum_X P_X(\tau) x_i Y_X(t-\tau-2) = O(\tau(t-\tau)^2\cdot r^\tau),$$

получаем, что

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}c_0\cdot(1-r)r^{t-\tau-2}(1+o(1))}{c_0 btu'_1(1-r)r^{t-1}} \times \Big[\sum_X P_X(\tau)x_i S''_{ij}$$

$$+\sum_X P_X(\tau)\cdot\sum_{l=1}^{k_1} u'_i x_i x_l \cdot br\cdot(t-\tau) + r\cdot\sum_X P_X(\tau)\cdot\sum_{l=k_1+1}^{k} u''_{l-k_1} x_i x'_l\Big]$$

$$= \frac{p_{ij}(1+o(1))}{u'_1 tr^{\tau+1}} \times \left(a_i^1(\tau)\cdot S''_{ij} + br(t-\tau)\cdot\sum_{l=1}^{k_1} u'_l b_{il}^1(\tau) + r\cdot\sum_{l=k_1+1}^{k} u''_{l-k_1} b_{il}^1(\tau)\right).$$

Используя следствие III.2.1 и лемму III.5.1 для первых и вторых моментов, имеем:

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}\cdot(1+o(1))}{t}\cdot\Big(\tau t(v''_{i-k_1}S''_{ij}/r + G''_i) + (t-\tau)\cdot G'_i\Big),$$

где $G''_i = \sum\limits_{l=k_1+1}^{k} u''_{l-k_1} g_{il}$. Полученные результаты сформулируем в виде теоремы.

Теорема III.5.1 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1), при $r = r' = r'' < 1$ и $\tau, t-\tau \to \infty$ справедливы равенства:*

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}\cdot(t-\tau)}{t}\cdot\left(G'_i + \frac{v'_i S'_{ij}}{r}\right)(1+o(1))$$

для $i \le k_1$,

$$M(q_{ij}(t,\tau)) = \frac{p_{ij}}{t}\cdot\big(\tau\cdot(v''_{i-k_1}S''_{ij}/r + G''_i) + (t-\tau)\cdot G'_i\big)(1+o(1))$$

для $k_1 < i \le k$. Здесь $j = 1,\ldots,n_i$, где n_i - количество правил грамматики с нетерминалом A_i в левой части. Величины $M(q_{ij}(t,\tau))$ не меняются при замене аксиомы на любой нетерминал из класса K_1.

Пользуясь этим результатом, можно получить асимптотику математического ожидания числа нетерминалов x_i $(i = 1,\ldots,k)$ на ярусе τ в деревьях вывода высоты t при $\tau\to\infty$, $t-\tau\to\infty$. Обозначим это математическое ожидание через $M(x_i(t,\tau))$. Очевидно, что $M(x_i(t,\tau)) = \sum\limits_{j=1}^{n_i} M(q_{ij}(t,\tau))$. Учитывая, что $\sum\limits_j p_{ij} = 1$, $\sum\limits_j p_{ij}s_{ij}^l = a_l^i$, имеем для $i \le k_1$:

$$\sum_j p_{ij}S'_{ij} = \sum_{l=1}^{k_1} u'_l a_l^i = ru'_i,$$

поскольку u' является правым собственным вектором матрицы $A^{(1)}$, и

$$\sum_j p_{ij} S''_{ij} = \sum_{l=k_1+1}^{k} u''_{l-k_1} a^i_l = r u''_{i-k_1},$$

при $i > k_1$, так как u'' является правым собственным вектором матрицы $A^{(3)}$.

Поэтому для $i \leq k_1$ можно записать:

$$M(x_i(t,\tau)) \sim \sum_j p_{ij} G'_i \cdot \frac{t-\tau}{t} + v'_i \sum_j p_{ij} S'_{ij} \cdot \frac{t-\tau}{r \cdot t} = \frac{t-\tau}{t} \cdot (v'_i u^a_i + G'_i).$$

Аналогично для $k_1 < i \leq k$ получаем

$$M(x_i(t,\tau)) \sim \sum_j p_{ij} G'_i \cdot \frac{t-\tau}{t} + \sum_j p_{ij} G''_i \cdot \frac{\tau}{t} + v''_{i-k_1} \cdot \sum_j p_{ij} S''_{ij} \cdot \frac{\tau}{tr}$$

$$= G'_i \cdot \frac{t-\tau}{t} + (G''_i + v''_{i-k_1} u''_{i-k_1}) \cdot \frac{\tau}{t}.$$

Таким образом, справедливо

Следствие III.5.1 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) с $r = r' = r'' < 1$ при $\tau, t-\tau \to \infty$ верны асимптотические равенства:*

$$M(x_i(t,\tau)) \sim (v'_i u'_i + G'_i) \cdot \frac{t-\tau}{t}$$

для $i \leq k_1$,

$$M(x_i(t,\tau)) \sim G'_i \cdot \frac{t-\tau}{t} + (G''_i + v''_{i-k_1} u''_{i-k_1}) \cdot \frac{\tau}{t}$$

для $k_1 < i \leq k$.

Также легко получить математическое ожидание $M(\omega(t,\tau))$ общего числа нетерминалов на ярусе τ в деревьях вывода высоты t:

Следствие III.5.2 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) с $r = r' = r'' < 1$ при $\tau, t-\tau \to \infty$ справедливо асимптотическое равенство:*

$$M(\omega(t,\tau)) \sim 1 + \sum_{i=1}^{k} G'_i \cdot \frac{t-\tau}{t} + \sum_{i=k_1+1}^{k} G''_i \cdot \frac{\tau}{t}.$$

Заметим, что $M(\omega(t,\tau))$ всегда больше 1, поскольку если дерево имеет высоту t, то на любом ярусе τ, $1 \leq \tau < t$ есть хотя бы один нетерминал.

Таким образом, видим, что в случае кратного перронова корня для разложимой грамматики с двумя классами нетерминалов, так же как и в неразложимом случае, $M(q_{ij}(t,\tau))$ состоит из двух слагаемых. Первое равно $p_{ij} \cdot (t-\tau) v'_i S'_{ij} / rt$ при $i \leq k_1$, или $p_{ij} \tau \cdot v''_{i-k_1} S''_{ij} / rt$ при $i > k_1$ и учитывает соответственно количество нетерминалов из класса K_1 при $i \leq k_1$ или количество нетерминалов из класса K_2 при $i > k_1$ в правой части правила r_{ij}. Второе равно $p_{ij}\cdot(t-\tau)G'_i/t$ при $i \le k_1$, или $p_{ij}\cdot(\tau G''_i + (t-\tau)\cdot G'_i)\,/t$ при $i > k_1$ и определяется вторыми моментами грамматики. При этом, как и в неразложимом случае, при $t \to \infty$ происходит перераспределение частот правил : правила, порождающие больше нетерминальных символов, используются чаще в выводе слов, что, по-видимому, нужно для достижения большей высоты дерева вывода. Кроме того, $M(q_{ij}(t,\tau))$ ограничены сверху константами, поскольку отношения τ/t и $(t-\tau)/t$ не превосходят 1.

Возникают следующие отличия по сравнению с неразложимым случаем:
1) Величины $M(q_{ij}(t,\tau)),\ M(x_i(t,\tau)),\ M(\omega(t,\tau))$ зависят линейно от соотношения τ/t.
2) $M(q_{ij}(t,\tau))$ убывают для $i \le k_1$ при увеличении τ, а для $i > k_1$ могут как возрастать, так и убывать (см. пример раздела III.7).

Рассмотрим теперь величину $S_{ij}(t)/t$ - среднее число правил r_{ij} на один ярус дерева вывода высоты t. Введем константы $\omega_{ij}^{(m)},\ m=1,2,\ i,j=1,\ldots,k$ равенствами

$$\begin{array}{l} \omega_{ij}^{(1)} = p_{ij}\cdot\left(G'_i + v'_i S'_{ij}/r\right) \text{ при } i \le k_1, \\ \omega_{ij}^{(1)} = p_{ij} G'_i \text{ при } i > k_1, \\ \omega_{ij}^{(2)} = p_{ij}\cdot\left(G''_i + v''_{i-k_1} S''_{ij}/r\right) \text{ при } i > k_1, \end{array} \tag{III.5.5}$$

$$\begin{array}{l} \omega_{ij} = \omega_{ij}^{(1)}/2 \text{ при } i \le k_1, \\ \omega_{ij} = (\omega_{ij}^{(1)} + \omega_{ij}^{(2)})/2 \text{ при } i > k_1. \end{array} \tag{III.5.6}$$

Тогда можно записать, что

$$\begin{array}{l} M(q_{ij}(t,\tau)) = \omega_{ij}^{(1)}\cdot(t-\tau)/t \text{ при } i \le k_1, \\ M(q_{ij}(t,\tau)) = \omega_{ij}^{(1)}\cdot(t-\tau)/t + \omega_{ij}^{(2)}\cdot\tau/t \text{ при } i > k_1. \end{array}$$

Теорема III.5.2 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) при $r = r' = r'' < 1$ и $t \to \infty$ справедливо соотношение*

$$M(S_{ij}(t)/t) \to \omega_{ij},\ i = 1,\ldots,k,$$

где величины ω_{ij} определены формулами (III.5.6).

Доказательство. Теорема доказывается аналогично соответствующему утверждению для неразложимой грамматики в докритическом случае [9]. Рассмотрим случай $i \le k_1$. Положим $\tau_0 = \lfloor \log\log t \rfloor$. Представим $S_{ij}(t)$ в виде $S^1_{ij}(t)+S^2_{ij}(t)+S^3_{ij}(t)$, где

$$S^1_{ij}(t) = q_{ij}(t,1) + \ldots + q_{ij}(t,\tau_0),$$

$$S^2_{ij}(t) = q_{ij}(t,\tau_0+1) + \ldots + q_{ij}(t,t-\tau_0),$$

$$S^3_{ij}(t) = q_{ij}(t,t-\tau_0+1) + \ldots + q_{ij}(t,t).$$

В [9] (без предположений о неразложимости грамматики) была доказана оценка $M(S^1_{ij}(t)) \le \log^c(t)$, где $c = \log(k_{max}) + 1$, и k_{max} - максимальное количество нетерминалов в правой части правил грамматики. Также в [9] без предположений о неразложимости грамматики была получена следующая оценка для $S^3_{ij}(t)$:

$$S^3_{ij}(t) \leq \frac{a^1_i(t-\tau_0)}{P(D^t)} = O\left(\frac{1}{t \cdot r^{t-\tau_0}}\right) = \log^{c_1}(t)$$

для некоторой константы c_1.

Для $S^2_{ij}(t)$ получаем:

$$M(S^2_{ij}(t)) = \omega^{(1)}_{ij} \cdot \sum_{\tau=\tau_0}^{t-\tau_0} \left(\frac{t-\tau}{t}\right)(1+o(1))$$

$$= \omega^{(1)}_{ij} \cdot \sum_{\tau=1}^{t} \left(\frac{t-\tau}{t}\right)(1+o(1)) = \omega^{(1)}_{ij} \cdot t/2 + o(t).$$

Пользуясь неравенством

$$M(S^2_{ij}(t)) \leq M(S_{ij}(t)) \leq M(S^1_{ij}(t)) + M(S^2_{ij}(t)) + M(S^3_{ij}(t)),$$

находим, что

$$\frac{1}{t}\left(\frac{t\omega^{(1)}_{ij}}{2}\right)(1+o(1)) \leq M(S_{ij}(t)/t) \leq \frac{1}{t}\left(\frac{t\omega^{(1)}_{ij}}{2} + O(\log^{c_2}(t))\right)(1+o(1))$$

при $\tau_0 \to \infty$, где $c_2 = \max(c, c_1)$. Таким образом, устремляя τ_0 (а значит, и t) к бесконечности, получаем необходимое.

Для случая $i > k_1$ доказательство проводится аналогично. Единственное отличие заключается в оценке $S^2_{ij}(t)$:

$$M(S^2_{ij}(t)) = \omega^{(1)}_{ij} \cdot \sum_{\tau=\tau_0}^{t-\tau_0} \frac{t-\tau}{t}(1+o(1)) + \omega^{(2)}_{ij} \cdot \sum_{\tau=\tau_0}^{t-\tau_0} \frac{\tau}{t}(1+o(1))$$

$$= \omega^{(1)}_{ij} \cdot \sum_{\tau=1}^{t} \frac{t-\tau}{t}(1+o(1)) + \omega^{(2)}_{ij} \cdot \sum_{\tau=1}^{t} \frac{\tau}{t}(1+o(1)) = (\omega^{(1)}_{ij} + \omega^{(2)}_{ij}) \cdot t/2 + o(t).$$

Эта теорема показывает, что величина $M(S_{ij}(t)/t)$ стремится к константе при $t \to \infty$, как и в неразложимом случае.

III.6 Дисперсия математических ожиданий числа применения правил в деревьях вывода

Теперь оценим дисперсию величин $S_{ij}(t)/t$.

Теорема III.6.1 *Для КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) при $r = r' = r'' < 1$ и $t \to \infty$ справедливы соотношения*

$$D(S_{ij}(t)/t) \to \left(\omega_{ij}^{(1)}\right)^2 /12 \text{ при } i \le k_1,$$

$$D(S_{ij}(t)/t) \to \left(\omega_{ij}^{(2)} - \omega_{ij}^{(1)}\right)^2 /12 \text{ при } i > k_1,$$

где величины $\omega_{ij}^{(l)}$ определены формулами (III.5.5).

Доказательство этой теоремы использует тот же метод, что и доказательство аналогичной теоремы для неразложимой грамматики [9]. Очевидно,

$$D(S_{ij}(t)/t) = M\left((S_{ij}(t)/t)^2\right) - (\omega_{ij}(t))^2.$$

Вычислим $M\left(S_{ij}(t)^2\right)$:

$$M\left(S_{ij}(t)^2\right) = M\left(\left(\sum_{\tau=1}^{t} q_{ij}(t,u)\right)^2\right) = \sum_{\tau=1}^{t} M\left(q_{ij}(t,\tau)^2\right)$$

$$+2 \cdot \sum_{1 \le \tau_1 < \tau_2 \le t} M\left(q_{ij}(t,\tau_1) q_{ij}(t,\tau_2)\right).$$

Обозначим $S_1 = \sum\limits_{\tau=1}^{t} M\left(q_{ij}(t,\tau)^2\right)$ и $S_2 = 2 \cdot \sum\limits_{1 \le \tau_1 < \tau_2 \le t} M\left(q_{ij}(t,\tau_1) q_{ij}(t,\tau_2)\right)$. Пусть $l_* = \lfloor \log\log t \rfloor$. Разобьем сумму S_2 на семь частей в соответствии с тем, какие из величин $\tau_1, \tau_2 - \tau_1, t - \tau_2$ меньше l_* (все быть меньше l_* не могут). Буквально повторяя рассуждения [9] для нашего случая, получаем, что все части суммы S_2, кроме той, в которой $\tau_1, \tau_2 - \tau_1, t - \tau_2 \ge l_*$, имеют порядок $O(\log^{c_0} t)$ для некоторой константы c_0. Далее, сумма S_1 имеет порядок $O(t)$, что доказывается в точности так же, как соответствующее утверждение для неразложимого случая.

Вычислим главный член $S_2^* = \sum\limits_{l_* < \tau_1, \tau_1 + l_* < \tau_2 < t - l_*} M\left(q_{ij}(t,\tau_1) q_{ij}(t,\tau_2)\right)$. Пусть $d \in D^t$, а d_0 - поддерево d, имеющее высоту $\tau_1 + 1$. Количество нетерминалов A_i на ярусе $\tau_1 + 1$ в d и d_0 обозначим через $z_i, i = 1, \ldots, k$. Очевидно, если $q_{hl}(d_0)$ - число применений правила r_{hl} на ярусе τ_1, то $z_i = \sum\limits_{h=1}^{k} \sum\limits_{l=1}^{n_h} q_{hl}(d_0) s_{hl}^i, \ i = 1, \ldots, k$.

Зафиксируем поддерево d_0 и найдем вклад $S(d_0)$ в S_2^* деревьев из D^t, при удалении из которых вершин на ярусах с номерами, большими $\tau_1 + 1$, оставшееся поддерево совпадает с d_0. Обозначим вышеописанное подмножество деревьев из D^t через $T(d_0)$. Для любого дерева d из $T(d_0)$ условная вероятность d в D^t равна

$$p(d) = \frac{1}{P(D^t)} p(d_0) \prod_{i=1}^{z} p(d_0^i),$$

где $z = \sum_i z_i$, а d_0^i - поддерево d с корнем в i-й по порядку вершине на ярусе $\tau_1 + 1$. Обозначим через $q_{ij}^l(d)$ число применения правила r_{ij} в d_0^l, через $q_{ij}(\tau_1)$ - число применений правила r_{ij} на ярусе τ_1 в дереве d_0, а $x_i(\tau_1)$ - число нетерминалов A_i на ярусе τ_1 в дереве d_0. Очевидно, что

$$S(d_0) = \frac{1}{P(D^t)} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \sum_{d \in T(d_0)} p(d_0^1) \dots p(d_0^z) q_{ij}^l(d).$$

Рассмотрим сумму $\delta_l = \sum\limits_{d \in T(d_0)} p(d_0^1) \dots p(d_0^z) q_{ij}^l(d)$. Как и в [9], разобьем δ_l на две части : $\delta_l = \delta_l^{(1)} + \delta_l^{(2)}$, где $\delta_l^{(1)}$ учитывает деревья, в которых ярус t достигается через поддерево d_0^l, а $\delta_l^{(2)}$ – все остальные. В [9] доказано, что

$$\delta_l^{(1)} = Q_{Z'}(t - \tau_1 - 1) P(D_{m_l}^{t - \tau_1 - 1}) M(q_{ij}^l(t - \tau_1 - 1, \tau_2 - \tau_1 - 1)),$$

а

$$\delta_l^{(2)} < a_i^{m_l}(\tau_2 - \tau_1 - 1) R_{Z'}(t - \tau_1 - 1).$$

Заметим, что

$$S_2^* = \frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \delta_l^1 + \frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \delta_l^2.$$

Обозначим второе слагаемое через S_{22}^*. Очевидно, $z_i < c_0 x_i(\tau_1)$, где c_0-максимальное число нетерминалов в правых частях правил грамматики. Поэтому пользуясь найденными оценками для моментов грамматики, а также оценками для $Q_X(t)$ и $R_X(t)$, путем полного повторения соответствующего рассуждения для неразложимого случая, получаем, что

$$S_{22}^* = \frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \delta_l^2 = O\Big((t - \tau_1) r^{\tau_2 - \tau_1} / t\Big).$$

Первое слагаемое представим как

$$\frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \delta_l^1 = \frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} P_{m_l}(t - \tau_1 - 1) M(q_{ij}^l(t - \tau_1 - 1, \tau_2 - \tau_1 - 1)$$

$$+ \frac{1}{P(D^t)} \sum_{d} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} \left(Q_{Z'}(t - \tau_1 - 1)\right) P_{m_l}(t - \tau_1 - 1) M(q_{ij}^l(t - \tau_1 - 1, \tau_2 - \tau_1 - 1))$$

Обозначим вторую сумму через S_{23}^*. Оно выражает погрешность при замене $Q_{Z'}(t - \tau_1 - 1)$ на 1 в δ_l^1 и может быть оценено с использованием формул (III.4.1) и оценок для моментов грамматики как

$$S_{23}^* = \delta_l^{(1)} \cdot (1 - Q_{Z'}) = O\Big((t - \tau_1 - 1) r^{\tau_2 - \tau_1} / t\Big).$$

Вычислим теперь первую сумму

$$S_{21}^* = \frac{1}{P(D^t)} \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{z} P_{m_l}(t - \tau_1 - 1) M(q_{ij}^l(t - \tau_1 - 1, \tau_2 - \tau_1 - 1)). \quad \text{(III.6.1)}$$

Напомним, что при $t-\tau_2, \tau_2-\tau_1, \tau_1 \to \infty$ справедливы оценки

$$P_i(t-\tau_1-1) = bc_0 u_i' \cdot (1-r)(t-\tau_1-1) r^{t-\tau_1-2} \cdot (1+o(1)), \; i \le k_1,$$

$$P_i(t-\tau_1-1) = c_0 u_{i-k_1}'' \cdot (1-r) t^{t-\tau_1-2} \cdot (1+o(1)), \; i > k_1,$$

$$M(q_{ij}^l(t-\tau_1-1, \tau_2-\tau_1-1)) = \frac{t-\tau_2}{t-\tau_1-1} \cdot \omega_{ij}^{(1)}(1+o(1)) \text{ при } l \le k_1, i \le k_1,$$

$$M(q_{ij}^l(t-\tau_1-1, \tau_2-\tau_1-1)) = \left(\frac{t-\tau_2}{t-\tau_1-1} \cdot \omega_{ij}^{(1)} + \frac{\tau_2-\tau_1-1}{t-\tau_1-1} \cdot \omega_{ij}^{(2)} \right) \cdot (1+o(1)) \text{ при } l \le k_1, i > k_1$$

$$M(q_{ij}^l(t-\tau_1-1, \tau_2-\tau_1-1)) = 0, \; M(q_{ij}^l(t-\tau_1-1, \tau_2-\tau_1-1)) = \omega_{ij}^{(2)} \cdot (1+o(1)) \text{ при } l > k_1, i > k_1.$$

Далее мы рассмотрим случаи $i \le k_1$ и $i > k_1$ отдельно.
1) Пусть $i \le k_1$. В этом случае, пользуясь оценками для $P_i(t), M(q_{ij}^l(t,\tau))$, получаем, что

$$S_{21}^* = \frac{bc_0 \cdot (1-r)(t-\tau_2) r^{t-\tau_1-2} \omega_{ij}^{(1)}(1+o(1))}{bc_0 u_1'(1-r) t r^{t-1}} \cdot \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{k_1} u_l' z_l$$

$$= \frac{\omega_{ij}^{(1)} \cdot (t-\tau_2)(1+o(1))}{u_1' t r^{\tau_1+1}} \cdot \sum_{d_0} p(d_0) q_{ij}(\tau_1) \sum_{l=1}^{k_1} u_l' z_l$$

$$= \frac{\omega_{ij}^{(1)} \cdot (t-\tau_2)(1+o(1))}{u_1' t r^{\tau_1+1}} \cdot \sum_{d} p(d_0) q_{ij}(\tau_1) \sum_{l \le k_1} \sum_{m \le k_1} \sum_{n} u_l' s_{mn}^l q_{mn}(\tau_1). \qquad \text{(III.6.2)}$$

Найдем сумму $S_{ijhl} = \sum_d p(d_0) q_{ij}(\tau_1) q_{hl}(\tau_1)$. Для этого введем случайные величины $I_{ij}^m(d_0)$, которые равны единице, если в вершине с номером m, $m = 1, \ldots, x_i(\tau_1)$ среди помеченных нетерминалом A_i на ярусе τ_1 применено правило r_{ij}, и нулю в противном случае. Заметим, что

$$\sum_d p(d_0) I_{ij}^m(d_0) I_{hl}^n(d_0) = \begin{cases} P_X(\tau_1) p_{ij} p_{hl}, & n \ne m \\ P_X(\tau_1) p_{ij}, & m = n, i = h, j = l \\ 0, & \text{иначе,} \end{cases}$$

где $X = (x_1(\tau_1), \ldots, x_k(\tau_1))$ - вектор нетерминалов на ярусе τ_1. Представляя $q_{ij}(\tau_1)$ как $\sum\limits_{m=1}^{x_i(\tau_1)} I_{ij}^m(d_0)$, получаем, что

$$S_{ijhl} = \sum_d p(d_0) \sum_n \sum_m I_{ij}^n(d) I_{hl}^m(d)$$

$$= \begin{cases} p_{ij} p_{hl} \sum\limits_X P_X(\tau_1) x_i x_h = p_{ij} p_{hl} b_{ih}^1(\ddot{a}u_1), & i \ne h \\ p_{ij} p_{il} \sum\limits_X P_X(\tau_1) x_i(x_i-1) = p_{ij} p_{il} b_{ii}^1(\tau_1), & i = h, j \ne l \\ p_{ij}^2 \sum\limits_X P_X(\tau_1) x_i(x_i-1) + p_{ij} \sum\limits_X P_X(\tau_1) x_i = p_{ij}^2 b_{ii}^1(\tau_1) + p_{ij} a_i^1(\tau_1), & i = h, j = l. \end{cases}$$

(III.6.3)

Подставляя эти выражения в формулу (III.6.2), получаем:

$$S_{21}^{*}=\frac{p_{ij}\omega_{ij}^{(1)}\cdot(t-\tau_2)(1+o(1))}{u_1'tr^{\tau_1+1}}\cdot\left(\sum_{l,m\le k_1}u_l's_{mn}^l\sum_n p_{mn}b_{im}^1(\tau_1)+\sum_{l\le k_1}u_l's_{ij}^la_i^1(\tau_1)\right).$$

Так как

$$\sum_{l,m\le k_1}u_l'b_{im}^1(\tau_1)\sum_n p_{mn}s_{mn}^l=u_1'\cdot\sum_{l,m\le k_1}u_l'a_m^lg_{im}r^{\tau_1}\cdot(1+o(1))$$

$$=u_1'r^{\tau_1+1}\cdot\sum_m u_m'g_{im}\cdot(1+o(1))=u_1'r^{\tau_1+1}G_i'\cdot(1+o(1)),$$

и

$$\sum_{l\le k_1}u_l's_{ij}^la_i^1(\tau_1)=u_1'v_i'S_{ij}'r^{\tau_1}\cdot(1+o(1)),$$

то $S_{21}^{*}=\left(\omega_{ij}^{(1)}\right)^2\cdot\frac{t-\tau_2}{t}(1+o(1))$.

2) Пусть $i>k_1$. Подставляя аналогичным образом оценки для $P_i(t),M(q_{ij}^l(t,\tau))$ в формулу (III.6.1), получаем

$$S_{21}^{*}=\frac{bc_0\cdot(1-r)r^{t-\tau_1-2}(1+o(1))}{bu_1'c_0\cdot(1-r)tr^{t-1}}\cdot\left((\tau_2-\tau_1-1)\omega_{ij}^{(1)}+(t-\tau_2)\omega_{ij}^{(2)}\right)$$

$$\times\sum_{d_0}p(d_0)q_{ij}(\tau_1)\sum_{l\le k_1}u_l'z_l+\frac{c_0(1-r)r^{t-\tau_1-1}\omega_{ij}^{(2)}\cdot(1+o(1))}{bu_1'c_0(1-r)tr^{t-1}}\cdot\sum_{d_0}p(d_0)q_{ij}(\tau_1)\sum_{l>k_1}u_{l-k_1}''z_l.$$

Пользуясь формулами (III.6.3) и учитывая, что $z_l=\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i}s_{ij}^lq_{ij}(\tau_1)$, аналогично находим, что

$$\sum_{d_0}p(d_0)q_{ij}(\tau_1)\sum_{l\le k_1}u_l'z_l=\sum_{d_0}q_{ij}(\tau_1)\sum_{l,m\le k_1}\sum_n u_l's_{mn}^lq_{mn}(\tau_1)$$

$$=p_{ij}\sum_{d_0}q_{ij}(\tau_1)\cdot\sum_{l,m}u_l'a_m^lb_{im}^1(\ddot{a}u_1)=\omega_{ij}^{(1)}r^{\tau_1+1}(1+o(1)),$$

$$\sum_{d_0}p(d_0)q_{ij}(\tau_1)\sum_{l>k_1}u_{l-k_1}''z_l=\sum_{d_0}q_{ij}(\tau_1)\sum_{l,m>k_1,n}u_{l-k_1}''s_{mn}^lq_{mn}(\tau_1)$$

$$+\sum_{d_0}q_{ij}(\tau_1)\sum_{l\le k_1,m>k_1}\sum_n u_{l-k_1}''s_{mn}^lq_{mn}(\tau_1)=u_1'p_{ij}r^{\tau_1}\cdot(1+o(1))\cdot\big(\tau_1bv_{i-k_1}''S_{ij}''$$

$$+\tau_1brG_{i-k_1}''+G_{i-k_1}'\big)=\omega_{ij}^{(2)}b\tau_1r^{\tau_1}(1+o(1)).$$

Следовательно, при $i>k_1$

$$S_{21}^{*}\sim\left(\omega_{ij}^{(1)}\right)^2\cdot\frac{\tau_2-\tau_1-1}{t}+\omega_{ij}^{(1)}\omega_{ij}^{(2)}\cdot\frac{t-\tau_2}{t}+\left(\omega_{ij}^{(2)}\right)^2\cdot\frac{\tau_1}{t}.$$

Пользуясь полученными выражениями для S^*_{21}, видим, что погрешностью, возникающей при замене $Q_{Z'}$ на 1 и отбрасывании δ^2_l, можно пренебречь, так как в сумме $S^*_2 = S^*_{21} + S^*_{22} + S^*_{23}$ член $S^*_{22} + S^*_{23} = o(S^*_{21})$ в силу доказанных выше оценок для S^*_{22}, S^*_{23}.

Таким образом, при $i \le k_1$

$$S_2 = 2 \cdot \sum_{l_* < \tau_1, \tau_1 + l_* < \tau_2 < t - l_*} M\Big(q_{ij}(\tau_1) q_{ij}(\tau_2)\Big) + O(\log^c(t)) =$$

$$\sum_{l_* < \tau_1, \tau_1 + l_* < \tau_2 < t - l_*} \left(\omega^{(1)}_{ij}\right)^2 (2 + o(1)) \cdot \frac{t - \tau_2}{t} + O(\log^c(t)) =$$

$$\left(\omega^{(1)}_{ij}\right)^2 (2 + o(1)) \cdot \sum_{1 < \tau_1 < \tau_2 < t} \frac{t - \tau_2}{t} = \left(\omega^{(1)}_{ij} t\right)^2 (1 + o(1))/3.$$

Аналогично, при $i > k_1$

$$S_2 = 2 \cdot \left(\sum_{1 < \tau_1 < \tau_2 < t} \left(\omega^{(1)}_{ij}\right)^2 \cdot \frac{\tau_2 - \tau_1 - 1}{t} + \omega^{(1)}_{ij} \omega^{(2)}_{ij} \cdot \frac{t - \tau_2}{t} + \left(\omega^{(2)}_{ij}\right)^2 \cdot \frac{\tau_1}{t} \right) \cdot (1 + o(1)) + O(\log^c(t))$$

$$= \frac{t^2 \cdot (1 + o(1))}{3} \left(\left(\omega^{(1)}_{ij}\right)^2 + \left(\omega^{(2)}_{ij}\right)^2 + \omega^{(1)}_{ij} \omega^{(2)}_{ij} \right).$$

Учитывая, что $S_1 = O(t)$, получаем, что при $i \le k_1$

$$D(S_{1j}(t)/t) \to \left(\omega^{(1)}_{ij}\right)^2 /3 - \left(\omega^{(1)}_{ij}/2\right)^2 = \left(\omega^{(1)}_{ij}\right)^2 /12,$$

$$D(S_{2j}(t)/t) \to \left(\left(\omega^{(1)}_{ij}\right)^2 + \left(\omega^{(2)}_{ij}\right)^2 + \omega^{(1)}_{ij} \omega^{(2)}_{ij} \right) /3 - \left((\omega^{(1)}_{ij} + \omega^{(2)}_{ij})/2 \right)^2 = \left(\omega^{(2)}_{ij} - \omega^{(1)}_{ij} \right)^2 /12.$$

Теорема доказана.

Таким образом, видно, что в разложимом случае, в отличие от неразложимого, дисперсия количества правил r_{ij} на один ярус дерева вывода отлична от нуля, и соответственно, слова не обладают почти постоянным побуквенным составом.

III.7 Пример грамматики с двумя классами нетерминалов

Приведем пример применения формул для математических ожиданий. Рассмотрим грамматику с двумя нетерминалами A_1, A_2, и $k_1 + k_3$ терминалами $y_i, i = 1, \dots, k_1,\ x_i, i = 1, \dots, k_3$ и следующими правилами :

$$r_{11} : A_1 \overset{p/k_1}{\rightarrow} A_1 y_1 A_1 y_2 \dots A_1 y_{k_1} A_2^{k_2},$$

$$r_{12} : A_1 \overset{1-p/k_1}{\rightarrow} \lambda,$$

$$r_{21} : A_2 \overset{p/k_3}{\to} A_2 x_1 \dots A_2 x_{k_3},$$

$$r_{22} : A_2 \overset{1-p/k_3}{\to} \lambda,$$

где λ - пустое слово, $0 < p < 1$. К примеру, для $k_1 = 1, k_2 = k_3 = 2$ такая грамматика порождает слова вида $y_1^n V_1 \dots V_n$, где V_i - правильные скобочные выражения от x_1, x_2, если отождествить x_1 с открывающей скобкой, а x_2 с закрывающей. Для данной грамматики $K_1 = \{A_1\}, K_2 = \{A_2\}$. Матрица первых моментов имеет вид

$$A = \begin{pmatrix} p & pk_2/k_1 \\ 0 & p \end{pmatrix},$$

а вторые моменты представлены следующими выражениями:

$$b_{11}^1 = p(k_1 - 1),\ b_{12}^1 = pk_2,\ b_{22}^1 = pk_2(k_2 - 1)/k_1,\ b_{22}^2 = p(k_3 - 1).$$

Используя формулы (III.5.4) для констант g_{ij}, получаем:

$$g_{11}(t) = (k_1 - 1)/(1 - p),\ g_{22}(t) = (k_3 - 1)/(1 - p),\ g_{12}(t) = \frac{k_2}{1 - p} + \frac{pk_2(k_1 - 1)}{(1 - p)^2 k_1}.$$

Используя теорему III.5.1, находим, что

$$M(q_{11}(t, \tau)) \sim \frac{1 - p/k_1}{1 - p} \cdot \frac{t - \tau}{t},$$

$$M(q_{12}(t, \tau)) \sim \frac{(k_1 - p)(k_1 - 1)}{k_1(1 - p)} \cdot \frac{t - \tau}{t},$$

$$M(q_{21}(t, \tau)) \sim \frac{p}{k_3} \cdot \left[\frac{k_3 - p}{p(1 - p)} \frac{\tau}{t} + \frac{pk_2(k_1 - 1)}{k_1(1 - p)^2} \frac{t - \tau}{t}\right],$$

$$M(q_{22}(t, \tau)) \sim \left(1 - \frac{p}{k_3}\right) \cdot \left[\frac{k_3 - 1}{1 - p} \frac{\tau}{t} + \frac{pk_2(k_1 - 1)}{k_1(1 - p)^2} \frac{t - \tau}{t}\right]$$

при $\tau \to \infty, t - \tau \to \infty$.

Математическое ожидание числа применений правил r_{11} убывает с ростом τ. Для r_{12} оно убывает при $k_1 > 1$ и равно нулю при $k_1 = 1$ (в этом случае правило r_{12} применяется только один раз в выводе любого слова). Математическое ожидание числа применений правил r_{21}, r_{22} может возрастать, убывать или быть постоянным в зависимости от параметров k_1, k_2, k_3, p. Например, для $k_1 = 1, k_2 = k_3 = 2$ математическое ожидание числа применений правил r_{21} линейно возрастает, а r_{22} - асимптотически постоянно, а для $k_1 = 2, k_3 = 3, k_2 = 1, p = 0.5$ оба математических ожидания убывают.

III.8 Оценка стоимости оптимального кодирования

В этом разделе будем рассматривать только языки, порожденные КС-грамматиками с однозначным выводом. Сначала получим асимптотическую формулу для энтропии

$H(\mathcal{L}^t)$ множества слов $\mathcal{L}^t$. В дальнейшем мы рассматриваем только случай кратного перронова корня, поскольку иначе оценки для $H(\mathcal{L}^t)$ и $C^*(\mathcal{L})$, как легко заметить, имеют такой же вид, как и в неразложимом случае. Поэтому в дальнейшем предполагаем, что матрица первых моментов грамматики имеет вид (I.2.1) и $r = r' = r'' < 1$.

По определению, $H(\mathcal{L}^t) = -\sum\limits_{\alpha \in L^t} p_t(\alpha) \cdot \log p_t(\alpha)$. Следовательно,

$$H(\mathcal{L}^t) = -\sum_{\alpha \in L^t} p_t(\alpha) \left(\log p(\alpha) - \log P(L^t) \right)$$

$$= \frac{1}{P(L^t)} \cdot \left(-\sum_{\alpha \in L^t} p(\alpha) \log p(\alpha) \right) + \log P(L^t).$$

Для слова α обозначим через $q_{ij}(\alpha)$ число применений правила r_{ij} при его выводе. Вероятность слова α равна $p(\alpha) = \prod\limits_{i=1}^{k} \prod\limits_{j=1}^{n_i} p_{ij}^{q_{ij}(\alpha)}$, соответственно, $\log p(\alpha) = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n_i} q_{ij}(\alpha) \cdot \log p_{ij}$. Поэтому можно записать, что

$$H(\mathcal{L}^t) = \frac{1}{P(L^t)} \cdot \left(-\sum_{\alpha \in L^t} p(\alpha) \cdot \sum_{i=1}^{k} \sum_{j=1}^{n_i} q_{ij}(\alpha) \cdot \log p_{ij} \right) + \log P(L^t)$$

$$= \frac{1}{P(L^t)} \cdot \left(-\sum_{i=1}^{k} \sum_{j=1}^{n_i} \log p_{ij} \cdot \sum_{\alpha \in L^t} p(\alpha) \cdot q_{ij}(\alpha) \right) + \log P(L^t).$$

Очевидно, что $\sum\limits_{\alpha \in L^t} p(\alpha) \cdot q_{ij}(\alpha) = P(L^t) \cdot M(S_{ij}(t))$. По теореме III.5.1 имеем $M(S_{ij}(t)) = t \cdot \omega_{ij} + o(t)$, где ω_{ij} введены формулами (III.5.6). Поэтому выражение для энтропии можно переписать в виде

$$H(\mathcal{L}^t) = -t \cdot (1 + o(1)) \cdot \sum_{i=1}^{k} \sum_{j=1}^{n_i} \omega_{ij} \log p_{ij} + \log P(L^t).$$

Ввиду однозначности вывода $\log P(L^t) = \log P(D^t) = t \cdot \log r + O(\log t)$, следовательно,

$$H(\mathcal{L}^t) = t \cdot \left(\log r - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \omega_{ij} \log p_{ij} \right) + o(t). \tag{III.8.1}$$

Полученный результат сформулируем в виде теоремы.

Теорема III.8.1 *Энтропия для множества слов с деревьями вывода высоты t из стохастического КС-языка, порожденного разложимой КС-грамматикой с матрицей первых моментов вида (I.2.1) при $r = r' = r'' < 1$ имеет вид*

$$H(\mathcal{L}^t) = t \cdot \left(\log r - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \omega_{ij} \log p_{ij} \right) + o(t).$$

Здесь величины ω_{ij} определены формулами (III.5.6), а p_{ij} - вероятность применения правила r_{ij}.

Таким образом, энтропия $H(\mathcal{L}^t)$ зависит асимптотически линейно от высоты дерева t, как и в неразложимом случае.

Оценим теперь стоимость оптимального кодирования для слов КС-языка, порожденного рассматриваемой грамматикой. Обозначим кодирование множества слов $\mathcal{L}^t$, минимизирующее величину $M_t(f) = \sum_{\alpha \in L^t} p_t(\alpha) \cdot |f(\alpha)|$, через f^*. Тогда очевидно, что для любого кодирования f множества слов $L^t \subset L$ справедливо неравенство $M_t(f) \geq M_t(f^*)$. Величину $M_t(f^*) = M^*(\mathcal{L}^t)$ оценим, используя теорему II.3. Поскольку $H(\mathcal{L}^t) \to \infty$ при $t \to \infty$, то из этой теоремы следует, что $M_t(f^*)/H(\mathcal{L}^t) \to 1$ при $t \to \infty$.

Теперь вычислим величину $\sum_{\alpha \in L^t} p_t(\alpha) \cdot |\alpha|$. Пусть правило r_{ij} содержит в правой части l_{ij} терминальных символов. Очевидно, что $|\alpha| = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n_i} q_{ij}(\alpha) l_{ij}$. Поэтому

$$\sum_{\alpha \in L^t} p_t(\alpha) \cdot |\alpha| = \sum_{i=1}^{k} \sum_{j=1}^{n_i} l_{ij} M(S_{ij}(t)) = t \cdot \sum_{i,j} l_{ij} \omega_{ij} + o(t).$$

Следовательно, справедлива

Теорема III.8.2 *Пусть $\mathcal{L}$ - КС-язык, порожденный разложимой грамматикой с матрицей первых моментов вида (I.2.1) с $r = r' = r'' < 1$. Тогда стоимость любого кодирования f этого языка удовлетворяет неравенству*

$$C(\mathcal{L}, f) \geq C^*(\mathcal{L}) = \frac{\log r - \sum_{i,j} \omega_{ij} \log p_{ij}}{\sum_{i,j} l_{ij} \omega_{ij}},$$

где ω_{ij} определены формулами (III.5.6), p_{ij} - вероятность применения правила r_{ij}, а l_{ij} - число терминальных символов в правой части правила r_{ij}.

III.9 Алгоритм асимптотически оптимального кодирования

Можно показать, что оценка, полученная в предыдущем параграфе, является точной, т. е. существует кодирование множества слов $\mathcal{L}^t$, стоимость которого сколь угодно близка к $C^*(\mathcal{L})$. Для этого будем использовать схему кодирования, предложенную в [11]. Как доказано выше, частота правила r_{ij} в деревьях вывода высоты $t \to \infty$ стремится к $p^*_{ij} = \omega_{ij}/\omega_i$, где $\omega_i = \sum_j \omega_{ij}$. Для каждого множества правил R_i с одинаковой левой частью A_i, применяется схема двоичного префиксного кодирования Шеннона. При этом правилу r_{ij} будет соответствовать элементарный код длины $\lceil -\log p^*_{ij} \rceil$ (здесь и далее будем обозначать через $\lceil x \rceil$ ближайшее целое к x сверху).

Код слова $\alpha \in L^t$, имеющего левый вывод $\omega(\alpha) = r_{i_1 j_1} r_{i_2 j_2} \cdots r_{i_m j_m}$, получается конкатенацией кодов правил $r_{i_s j_s}$. Обозначим такое кодирование f_{sh}. По аналогии с

[11] мы докажем, что стоимость кодирования f_{sh} стремится к $C^*(\mathcal{L})$ при переходе к укрупненной грамматике $G(n), n \to \infty$, что в некотором смысле эквивалентно блочному кодированию, рассмотренному Шенноном, причем блоку соответствует поддерево высоты n дерева вывода.

Обозначим величину $\sum_{i,j} l_{ij}\omega_{ij}$ за h. Стоимость кодирования f_{sh} равна

$$C(\mathcal{L}, f_{sh}) = \lim_{t\to\infty} \frac{\sum_{\alpha\in L^t} p_t(\alpha)\cdot|f_{sh}(\alpha)|}{th + o(t)}.$$

Числитель

$$\sum_{\alpha\in l^t} p_t(\alpha)\cdot|f_{sh}(\alpha)| = M\left(\sum_{i,j}\omega_{ij}(\alpha)\cdot\lceil -\log p^*_{ij}\rceil\right)$$

$$= t\cdot\sum_{i,j}\omega_{ij}\cdot\lceil -\log p^*_{ij}\rceil + o(t) = t\cdot\sum_{i,j}\omega_{ij}\cdot\lceil -\log(\omega_{ij}/\omega_i)\rceil + o(t).$$

Поэтому

$$C(\mathcal{L}, f_{sh}) = \frac{1}{h}\sum_{i,j}\omega_{ij}\cdot\lceil -\log(\omega_{ij}/\omega_i)\rceil.$$

Избыточность кодирования f_{sh} равна

$$C(\mathcal{L}, f_{sh}) - C^*(\mathcal{L}) = \frac{1}{h}\sum_{i,j}\omega_{ij}\lceil -\log p^*_{ij}\rceil - \frac{\log r}{h} + \frac{1}{h}\sum_{i,j}\omega_{ij}p_{ij}$$

$$\le \frac{1}{h}\sum_{i,j}\omega_{ij}\cdot\left(-\log p^*_{ij}+1\right) - \frac{\log r}{h} + \frac{1}{h}\sum_{i,j}\omega_{ij}p_{ij} = \frac{\omega}{h} - \frac{\log r}{h} + \frac{1}{h}\sum_{i,j}\log\frac{\omega_i p_{ij}}{\omega_{ij}}.$$

Рассмотрим укрупнение исходной грамматики $G(n)$ при $n \to \infty$. Рассмотрим множество слов L^t в грамматике $G(n)$, причем будем считать, что $t/n \to \infty$. Обозначим $t' = t/n$.

Поскольку матрица первых моментов грамматики $G(n)$ равна A^n, справедливо равенство $r' = r^n$, и собственные вектора u', v', u'', v'' не меняются при переходе к грамматике $G(n)$. Кроме того, справедливы равенства

$$\omega_i = \sum_j \omega_{ij} = \omega'_i = \sum_j \omega'_{ij},\ \omega = \sum_i \omega_i = \omega' = \sum_i \omega'_i.$$

Пусть

$$h' = \sum_{ij} l'_{ij}\omega'_{ij},\ C^*(\mathcal{L})' = \sum_{i,j}\omega'_{ij}\cdot\lceil -\log(\omega'_{ij}/\omega'_i)\rceil/h' + o(1),$$

где величины h', l'_{ij}, ω'_i соответствуют грамматике $G(n)$. Очевидно, что $h' = hn$.

Докажем, что избыточность кодирования f'_{sh}, полученного применением вышеописанного алгоритма кодирования к правилам грамматики $G(n)$, стремится к нулю при условиях $n \to \infty, t/n \to \infty$. Для этого нам понадобится несколько вспомогательных результатов. Обозначим через $L'^{t'}$ множество слов, имеющих деревья вывода высоты t' в новой грамматике.

Лемма III.9.1 *При $t \to \infty$ математическое ожидание $M*_t(S'_{ij})$ числа применений правила r'_{ij} укрупненной грамматики в деревьях вывода слов из L^t равно $t'\omega'_{ij} \cdot (1+o(1))$, где величины ω'_{ij} введены формулами (III.5.5) и соответствуют новой грамматике $G(n)$. Иными словами, математические ожидания числа применений правил новой грамматики в деревьях вывода для слов из L^t почти такие же, как и в деревьях вывода слов из $L'^{t'}$.*

Пусть d – дерево вывода некоторого слова из L^t в грамматике $G(n)$. Разобьем его на два "слоя": d^1 является поддеревом d, содержащим все вершины расположенные на ярусах $1, \ldots, t'-1$, а d^2 является набором поддеревьев d высоты 1, корни которых расположены на ярусе $t'-1$. По определению, $p(d) = p(d^1)p(d^2)$. Обозначим $D^1_X(t') = D'^{t'}_X(t'-1)$, и $D^2_X(t', t) = D^1_X$ - множество наборов поддеревьев d^2 высоты 1, имеющих вектор X нетерминалов на ярусе $t'-1$, где максимальная высота дерева в наборе равна $t - t'n$. Обозначив через $S'_{ij}(d^1)$, $S'_{ij}(d^2)$ соответственно число применений правила r'_{ij} в d_1 и d_2, можем записать:

$$M *_t (S'_{ij}) = \sum_X \sum_{d^1 \in D^1_X(t'), d^2 \in D^2_X(t', t)} p(d^1)p(d^2)\Big(S'_{ij}(d^1) + S'_{ij}(d^2)\Big)$$

$$= \sum_X \sum_{d^1} p(d^1)S'_{ij}(d^1) + \sum_X \sum_{d^1} p(d^1) \sum_{d^2} p(d^2)S'_{ij}(d^2).$$

Первая сумма равна $M(S'_{ij}(t'-1)) = (t'-1)\omega'_{ij} \cdot (1+o(1))$ по теореме III.5.2. Вторая сумма может быть оценена как

$$\sum_X \sum_{d^1} p(d^1) \sum_{d^2} p(d^2)S'_{ij}(d^2) \leq \sum_X \sum_{d^1} p(d^1) \sum_{d^2} p(d^2)x_i = \sum_X \sum_{d^1} p(d^1)x_i = a_i(t')/P(D^{t'}) = O($$

Отсюда и следует утверждение леммы.

Из леммы III.9.1 вытекает два следствия.

Следствие III.9.1 *Справедливо соотношение $C^*(\mathcal{L}) = C^*(\mathcal{L})'$.*

Это следствие означает, что значения стоимости оптимального кодирования, полученные по множествам L^t и $L'^{t'}$, совпадают.

Следствие III.9.2 *При $t \to \infty$ стоимость кодирования f'_{sh} множества слов L^t грамматики G равна*

$$C(\mathcal{L}, f'_{sh}) = \frac{1}{h'} \sum_{i,j} \omega'_{ij} \cdot \lceil -\log(\omega'_{ij}/\omega'_i) \rceil.$$

Это означает, что стоимость кодирования f'_{sh}, вычисленная по множеству слов L^t при $t \to \infty$, такая же, как и вычисленная по множеству слов $L'^{t'}$ при $t' \to \infty$.

Оба утверждения следуют из того, что энтропия множества слов L^t в новой грамматике и среднее число терминалов h' в дереве вывода слов из L^t определяется математическими ожиданиями числа применений правил r'_{ij} укрупненной грамматики в деревьях вывода слов из L^t.

Используя эти следствия, получаем, что избыточность кодирования f'_{sh} в грамматике $G(n)$ равна

$$\Delta(n) = C(\mathcal{L}, f_{sh}) - C^*(\mathcal{L}) = \frac{\omega}{h'} - \frac{\log r'}{h'} + \frac{1}{h'} \sum_{i,j} \log \frac{\omega_i p'_{ij}}{\omega'_{ij}}.$$

Преобразуя выражение для $\Delta(n)$, получаем:

$$\Delta(n) = \frac{\omega}{hn} - \frac{\log r}{h} + \frac{1}{hn} \sum_{i,j} \log \frac{\omega_i p'_{ij}}{\omega'_{ij}}.$$

Разобьем множество правил R'_i на два подмножества (по аналогии с заключительными и незаключительными правилами в неразложимом случае) : R^1_i - множество правил, содержащих хотя бы один нетерминал из того же класса, что и A_i, в правой части, и R^2_i - множество остальных правил. Обозначим в дальнейшем $u_i = u'_i, v_i = v'_i,\ B_i = G'_i, S_{ij} = S'_{ij}$ для $i \le k_1$, и $u_i = u''_i, v_i = v''_i,\ B_i = G'_i + G''_i, S_{ij} = S''_{ij}$ для $i > k_1$, где величины $G'_i,\ G''_i,\ S'_{ij},\ S''_{ij}, u'_i, u''_i, v'_i, v''_i$ соответствуют грамматике $G(n)$. Так как $\omega_i = u_i v_i + B_i$, то величины B_i также не меняются при переходе к грамматике $G(n)$.

Очевидно,

$$\Delta(n) = \frac{\omega}{hn} - \frac{\log r}{h} + \frac{1}{hn} \sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} \log \omega_i + \frac{1}{2hn} \sum_{i=1}^{k} \sum_{R^2_i} p'_{ij} B_i \log(2\omega_i / B_i)$$

$$+ \frac{1}{hn} \sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} \log \frac{2r'}{v_i S_{ij} + B_i r'}.$$

Заметим, что

$$\frac{1}{hn} \cdot \left(\omega + \sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} \log \omega_i + \sum_{i=1}^{k} \sum_{R^2_i} p'_{ij} B_i \log(2\omega_i / B_i)/2 \right) = O(1/n),$$

так как

$$\sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} \log \omega_i \le \sum_{i=1}^{k} \omega_i \log \omega_i,$$

$$\sum_{i=1}^{k} \sum_{R^2_i} p'_{ij} B_i \log(2\omega_i / B_i) \le 2k \log e \cdot \sum_{i} \omega_i.$$

(в последней оценке применено неравенство $\log x \ \le\ \log e \cdot x$). Данная оценка показывает, что влияние правил из R^2_i стремится к нулю при $n \to \infty$.

Таким образом,

$$\Delta(n) = \frac{\log r}{h} \cdot \left(\sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} - 1 \right) + \frac{1}{hn} \cdot \sum_{i=1}^{k} \sum_{R^1_i} \omega'_{ij} \log \frac{2}{v_i S_{ij} + B_i r'} + O(1/n).$$

Опять используя неравенство $\log x \le \log e \cdot x$, получаем, что

$$\sum_{i=1}^{k} \sum_{R_i^1} \omega'_{ij} \log \frac{2}{v_i S_{ij} + B_i r'} \le 2 \log e \cdot \sum_{i=1}^{k} \sum_{R_i^1} p'_{ij}/r' = O(1),$$

т.к. $\sum_{R_i^1} p'_{ij}$ - это вероятность деревьев вывода высоты не менее n в неразложимой грамматике, порожденной классом нетерминалов, содержащим A_i, с A_i в качестве аксиомы, и она равна $Q_i(n) = O(r^n)$, как доказано в [11]. Далее, заметим, что

$$\sum_{R_i^1} \omega'_{ij} = \omega_i - \sum_{R_i^2} \omega'_{ij} = \omega_i - B_i \cdot \sum_{R_i^2} p'_{ij}/2 = \omega_i - B_i/2 + O(r^n) = u_i v_i/2 + O(r^n).$$

Поэтому

$$\Delta(n) = \frac{\log r}{h} \cdot \left(\sum_{i=1}^{k} u_i v_i/2 - 1 + O(r^n) \right) + o(1) = o(1).$$

Доказательство завершено.

Часть IV

Закономерности в деревьях вывода слов и оптимальное кодирование. Критический случай

В данной главе мы рассмотрим разложимую грамматику с матрицей первых моментов вида (I.2.1), перронов корень r которой равен 1. Этот случай называется критическим по аналогии с теорией ветвящихся процессов. Будем пользоваться теми же обозначениями для матрицы первых моментов, их подматриц и собственных векторов, что и в третьей главе для докритического случая. Будем считать, что собственные вектора матрицы A удовлетворяют условиям и нормировке, указанным в лемме III.2.1.

Как и в докритическом случае, рассмотрим отдельно случай кратного ($r' = r''$) и некратного ($r' \ne r''$) перронова корня.

IV.1 Случай кратного перронова корня

В этом разделе мы рассмотрим случай $r' = r'' = 1$. Сначала выведем асимптотические формулы для вероятностей продолжения $Q_i(t)$, $i = 1, \ldots, k$, и вероятностей $P_i(t)$ деревьев вывода высоты t. Затем с помощью этих результатов найдем асимптотику для математических ожиданий количества применений правил r_{ij} в деревьях вывода. С помощью этих результатов будет найдена асимптотика энтропии множества слов языка с деревьями вывода фиксированной высоты и получена нижняя оценка для стоимости оптимального кодирования языка, порожденного грамматикой. Затем построим схему кодирования, стоимость которого равна полученной нижней оценке (такое кодирование, однако, не является эффективным). В конце раздела покажем, что эффективный алгоритм кодирования, предложенный для докритического случая, и здесь является асимптотически оптимальным.

IV.1.1 Вероятности продолжения

Выведем асимптотические формулы для вероятностей продолжения $Q_i(t),\ i = 1, \ldots, k$. Приведем одно необходимое утверждение из теории ветвящихся процессов [16].

Теорема IV.1.1.1 *Пусть* $F(s) = \sum\limits_{\alpha} P\{\xi = \alpha\} s^{\alpha}$*- вероятностная производящая функция случайного вектора* $\xi = \{\xi_1, \ldots, \xi_n\}$ *со значениями из* N^n*. Пусть* $d > 0$ *- целое число. Предположим, что все* β*-моменты* d*-го порядка* $m_\beta = M\xi^{[\beta]}$ *конечны. Тогда имеет место разложение*

$$F(s) = \sum_{\bar\beta < d} \frac{(-1)^{\bar\beta} m_\beta}{\beta!} (1 - s_1)^{\beta_1} \cdot \ldots \cdot (1 - s_n)^{\beta_n} + R(d, s),$$

где $\bar\beta = \beta_1 + \ldots + \beta_n$*, а остаточный член имеет вид*

$$R(d, s) = (-1)^d \sum_{\bar\beta = d} \varepsilon_\beta(s) (1 - s_1)^{\beta_1} \cdot \ldots \cdot (1 - s_n)^{\beta_n},$$

причем при $0 \le s \le s' \le 1$ *верны оценки*

$$0 \le \varepsilon_\beta(s) \le \varepsilon_\beta(s') \le \frac{m_\beta}{\beta!},\ \lim_{s \to 1-0} \varepsilon_\beta(s) = \frac{m_\beta}{\beta!}.$$

Все моменты ветвящегося процесса, соответствующего КС-грамматике, конечны, так как для КС-грамматики производящая функция $F(s_1, \ldots, s_k)$ является полиномом от $s_1, \ldots, s_k$. Применяя теорему IV.1.1.1 к производящей функции $F(s)$, получаем следующее разложение для $1 - F_i(s),\ i = 1, \ldots, k$, в окрестности 1:

$$1 - F_i(s_1, \ldots, s_k) = \sum_{j=1}^{k} a^i_j \cdot (1 - s_j) - \frac{1}{2} \cdot \sum_{j,l \le k_1} b^i_{jl} \cdot (1 - s_j)(1 - s_l) + R^1_i(d, s),\ i = 1, \ldots, k_1, \tag{IV.1.1.1}$$

$$1 - F_i(s_1, \ldots, s_k) = \sum_{j > k_1} a^i_j \cdot (1 - s_j) - \frac{1}{2} \cdot \sum_{j,l > k_1} b^i_{jl} \cdot (1 - s_j)(1 - s_l) + R^2_i(d, s),\ i = k_1 + 1, \ldots, k, \tag{IV.1.1.2}$$

где $d = 3, 4, \ldots,$ и $R^1_i(d, s), R^2_i(d, s)$ имеют вид

$$R^1_i(d, s) = -\frac{1}{2} \cdot \sum_{l > k_1 \text{или} j > k_1} b^i_{jl} \cdot (1 - s_j)(1 - s_l)$$

$$+\sum_{\bar\beta=3}^{d}\frac{(-1)^{\bar\beta}m^i_\beta}{\beta!}(1-s)^\beta+O\left(\sum_{\bar\beta=d+1}m^i_\beta\cdot(1-s)^\beta\right),\ i=1,\dots,k_1$$

$$R^2_i(d,s)=\sum_{\bar\beta=3}^{d}\frac{(-1)^{\bar\beta}m^{i+k_1}_\beta}{\beta!}(1-s)^\beta+O\left(\sum_{\bar\beta=d+1}m^{i+k_1}_\beta\cdot(1-s)^\beta\right),\ i=1,\dots k_2.$$

Очевидно, $m^{i+k_1}_\beta=0$, если хотя бы одно $\beta_j>0$ при $j\le k_1$. Поэтому выражение для $R^2_i(d,s)$ можно переписать как

$$R^2_i(d,s)=\sum_{2<\bar\beta\le d,\ \beta(1:k_1)=0}\frac{(-1)^{\bar\beta}m^{i+k_1}_\beta}{\beta!}\cdot(1-s_{k_1+1})^{\beta_{k_1+1}}\dots(1-s_k)^{\beta_k}$$

$$+O\left(\sum_{\bar\beta=d+1,\ \beta(1:k_1)=0}m^{i+k_1}_\beta\cdot(1-s_{k_1+1})^{\beta_{k_1+1}}\dots(1-s_k)^{\beta_k}\right),\ i=1,\dots,k_2.$$

Обозначим

$$x_i(t)=1-F_i(t,0)=Q_i(t),\ i=1,\dots k_1,$$
$$y_i(t)=1-F_{i+k_1}(t,0)=Q_{i+k_1}(t),\ i=1,\dots k_2,$$
$$x(t)=(x_1(t),\dots,x_{k_1}(t)),\ y(t)=(y_1(t),\dots,y_{k_2}(t)).$$

Подставляя $s_i=x_i(t)$ при $i=1,\dots,k_1$, и $s_{i+k_1}=y_i(t)$ при $i=1,\dots,k_2$ в равенства (IV.1.1.1), (IV.1.1.2), и пользуясь равенством $F(t+1,s)=F(F(t,s))$, получаем соотношения

$$x_i(t+1)=\sum_{j\le k_1}a^i_jx_j(t)+\sum_{j>k_1}a^i_jy_{j-k_1}(t)$$
$$-\frac{1}{2}\cdot\sum_{j,l\le k_1}b^i_{jl}x_j(t)x_l(t)+R^1_i(d,x(t),y(t)),\ i=1,\dots,k_1, \tag{IV.1.1.3}$$

$$y_{i-k_1}(t+1)=\sum_{j>k_1}a^i_jy_{j-k_1}(t)$$
$$-\frac{1}{2}\cdot\sum_{j,l>k_1}b^i_{jl}y_{j-k_1}(t)y_{l-k_1}(t)+R^2_i(d,y(t)),\ i=k_1+1,\dots,k, \tag{IV.1.1.4}$$

где $d=3,4,\dots$. Здесь записано $R^2_i(d,x(t),y(t))=R^2_i(d,y(t))$, поскольку этот член на самом деле не зависит от $x(t)$.

Из ранее выписанной оценки для $R^1_i(d,s)$ следует, что в выражении для $R^1_i(d,x(t),y(t))$ все члены не превосходят по модулю $O\Big(x_i(t)y_j(t)+y_i(t)y_j(t)+x_i(t)x_j(t)x_l(t)\Big)$.

Поскольку производящие функции $F_i(s)$, $i=k_1+1,\dots,k$, являются производящими функциями для грамматики, порожденной нетерминалами класса

K_2, то при $i > k_1$ формулы для вероятностей продолжения и вероятности деревьев вывода высоты t имеют вид [10]:

$$y_i(t) = Q_i(t) = \frac{2u''_{i-k_1} \cdot (1 + o(1))}{B_2 t},$$

$$y_i(t) - y_i(t+1) = P_i(t) = \frac{2u''_{i-k_1} \cdot (1 + o(1))}{B_2 t^2}, \qquad \text{(IV.1.1.5)}$$

где $i = k_1 + 1, \ldots, k$, и

$$B_2 = \sum_{i,l,m > k_1} v''_{i-k_1} b^i_{lm} u''_{l-k_1} u''_{m-k_1}. \qquad \text{(IV.1.1.6)}$$

Пусть $z(t), t = 1, 2, \ldots$ - некоторая последовательность векторов или скаляров. Обозначим через $\delta z(t)$ последовательность, члены которой имеют вид $\delta z(t) = z(t + 1) - z(t)$.

Лемма IV.1.1.1 *Пусть последовательность $z(t), t = 1, 2, \ldots$ удовлетворяет рекуррентному соотношению $\delta z(t) = f(t) - g(t) z(t)$, где при $t \to \infty$ выполняются условия*

$$g(t) \to 0, \;\; f(t)/g(t) \to 0, \;\; \sum_{i=1}^{t} g(i) \to \infty.$$

Пусть $g(t) > 0$ при любом $t > t_0$ для некоторого t_0. Тогда $z(t) \to 0$ при $t \to \infty$.

Возьмем некоторое число $C > 2$. Очевидно, можно считать без ограничения общности, что $0 < g(t) < 1/2$ для любого t (взяв элемент последовательности $z(t)$ с достаточно большим номером в качестве первого). Рассмотрим два случая. В первом случае для некоторого t_1 при всех $t_2 > t_1$ выполняется неравенство $|z(t_2)| \geq C \cdot |f(t_2)/g(t_2)|$. Во втором случае предполагаем обратное, т.е. для любого t_1 существует номер $t_2 > t_1$, такой что $|z(t_2)| < C \cdot |f(t_2)/g(t_2)|$. Докажем, что в обоих случаях утверждение леммы выполнено. Сначала рассмотрим первый случай. Заметим, что при $|z(t)| \geq C \cdot |f(t)/g(t)|$ из соотношения $\delta z(t) = f(t) - g(t) z(t)$ вытекают неравенства

$$|z(t+1) - z(t)| = |f(t) - g(t)z(t)| < |f(t)| + |g(t)z(t)| < g(t)|z(t)|(1 + 1/C),$$

$$\begin{aligned} &|z(t+1) - z(t)| = |f(t) - g(t)z(t)| \\ &\geq ||g(t)z(t)| - |f(t)|| = |g(t)z(t)| - |f(t)| > g(t)|z(t)|(1 - 1/C). \end{aligned}$$

При получении второй оценки было использовано неравенство $|z(t)g(t)| \geq C|f(t)| > |f(t)|$. Из этого же неравенства следует, что $z(t)(z(t+1) - z(t)) < 0$. Таким образом, $z(t+1)$ получается прибавлением к $z(t)$ противоположной по знаку величины $\delta z(t)$. Следовательно, справедливы неравенства

$$|z(t)| \cdot (1 - (C+1)g(t)/C) \leq |z(t+1)| \leq |z(t)| \cdot (1 - (C-1)g(t)/C). \qquad \text{(IV.1.1.7)}$$

Разделив эти неравенства на $|z(t)|$ и перемножив получившиеся оценки по $t = t_1, \dots, n-1$, получим неравенство

$$|z(t_1)| \cdot \prod_{t=t_1}^{n-1} \big(1 - (C+1)g(t)/C\big) \leq |z(n)| \leq |z(t_1)| \cdot \prod_{t=t_1}^{n-1} \big(1 - (C-1)g(t)/C\big),$$

т.е $|z(n)| \to 0$ при $n \to \infty$, поскольку по известному критерию [18] сходимости бесконечного произведения оба произведения сходятся к нулю при $n \to \infty$ ввиду того, что ряд $\sum\limits_{t=1}^{n} g(t)$ расходится и $g(t) > 0$.

Теперь рассмотрим второй случай. Зафиксируем произвольно малое $\varepsilon > 0$. Возьмем номер t_1 такой, что $g(t) < \varepsilon,\ |f(t)/g(t)| < \varepsilon$ при $t > t_1$, и соответствующий ему номер $t_2 > t_1$, такой что $|z(t_2)| < C \cdot |f(t_2)/g(t_2)|$. Тогда для любого $t > t_2$ справедливо одно из двух утверждений:
1) $|z(t)| < C \cdot |f(t)/g(t)| < C\varepsilon$;
2) Существует $t_3,\ t_2 \leq t_3 < t$, такое что $|z(n)| > C \cdot |f(n)/g(n)|$ при $n = t, t-1, \dots, t_3+1$, и $|z(t_3)| < C \cdot |f(t_3)/g(t_3)| < C\varepsilon$.
Поскольку при $|z(t)| > C \cdot |f(t)/g(t)|$ из неравенства (IV.1.1.7) следует, что $|z(t+1)| < |z(t)|$, то во втором случае $|z(t)| < |z(t_3+1)| < |z(t_3)| + |f(t_3)| < (C+1)\varepsilon$. Устремляя $\varepsilon \to 0$, получаем, что $|z(t)| \to 0$ при $t \to \infty$. Лемма доказана.

Для решения рекуррентных соотношений (IV.1.1.3) нам понадобится также следующая лемма.

Лемма IV.1.1.2 *Пусть последовательность* $\{x_t\}$, $x_t > 0$ *при любом* $t \geq 0$, *удовлетворяет рекуррентному соотношению*

$$x_{t+1} = at^{\alpha}(1 + \varepsilon_1(t)) + (1 - bt^{\beta} \cdot (1 + \varepsilon_2(t)))x_t, \tag{IV.1.1.8}$$

где $\beta < 0$, $b > 0$, *и* $\varepsilon_1(t), \varepsilon_2(t) = o(1)$ *при* $t \to \infty$. *Тогда верны следующие асимптотические равенства:*
(1) $x_t = \frac{at^{\alpha+1}}{\alpha+1}(1 + o(1))$ *при* $\beta < -1,\ \alpha \geq 0$,
(2) $x_t = \frac{at^{\alpha+1}}{\alpha+b+1}(1 + o(1))$ *при* $\beta = -1,\ \alpha > -1$,
(3) $x_t = \frac{at^{\alpha-\beta}}{b}(1 + o(1))$ *при* $-1 < \beta < 0$.

Доказательство. Сначала докажем (1). Обозначим $x_*(t) = \frac{at^{\alpha+1}}{\alpha+1}$. Просуммировав по $i = 0, \dots, t$ равенство (IV.1.1.8), получаем

$$x_{t+1} = x_0 + a \cdot \sum_{i=0}^{t} i^{\alpha} \cdot (1 + o(1)) - b \cdot \sum_{i=0}^{t} i^{\beta}(1 + o(1))x_i \leq$$

$$\leq a \cdot \int_0^t i^{\alpha} di \cdot (1 + o(1)) = \frac{at^{\alpha+1}}{\alpha+1} \cdot (1 + o(1)) = x_*(t) \cdot (1 + o(1)).$$

Пользуясь этой оценкой, получаем, что

$$\sum_{i=0}^{t} i^{\beta} x_i = O(t^{\alpha+\beta+2}).$$

Подставляя эту оценку в предыдущее равенство, получаем, что $x_{t+1} = x_*(t)\cdot(1+o(1))$.

Равенство (2) доказывается применением леммы IV.1.1.1 к величине $z_t = (x_t t^{-\alpha-1} - a/(\alpha+b+1))$. Действительно, подставив $x_t = (z_t + a/(\alpha+b+1)) \cdot t^{\alpha+1}$ в соотношение (IV.1.1.8) и разделив на $(t+1)^{\alpha+1}$, получим соотношение

$$\delta z_t = c_0 \varepsilon_3(t)/t - z_t \cdot (\alpha + b + 1 + \varepsilon_4(t))/t$$

для некоторой константы c_0, где $\varepsilon_3(t), \varepsilon_4(t) = o(1)$ при $t \to \infty$, из которого по лемме IV.1.1.1 следует, что $z(t) \to 0$.

Равенство (3) доказывается применением леммы IV.1.1.1 к величине $z_t = (x_t t^{\beta-\alpha} - a/b)$. Аналогичным образом подставив $x_t = (z_t + a/b)t^{\alpha-\beta}$ в (IV.1.1.8) и разделив на $t^{\alpha-\beta}$, получим соотношение

$$\delta z_t = c_0 \varepsilon_3(t) t^{\beta+1} - b z_t t^{\beta+1}(1 + \varepsilon_4(t))$$

для некоторой константы c_0 и $\varepsilon_3(t), \varepsilon_4(t) = o(1)$.

Введем следующие обозначения:

$$\begin{aligned} x_*(t) &= \sum_{i=1}^{k_1} v_i' x_i(t), \\ y_*(t) &= \sum_{i=1}^{k_2} v_i'' y_i(t). \end{aligned} \tag{IV.1.1.9}$$

В силу соотношений (IV.1.1.3), (IV.1.1.4) справедливо равенство

$$Q(t+1) = (x(t), y(t)) = (A - E_t) \cdot Q(t),$$

где $E_t = O(Q(t)) = O\Big(\max(x(t), y(t))\Big)$, причем $0 \le E_t \le A$ при достаточно большом t.

Пусть n_0 таково, что $E_t \ge 0$ при $t > n_0$. Тогда можно записать:

$$Q(n) = \prod_{t=n_0}^{n-1} (A - E_t) \cdot Q(n_0).$$

Поскольку вероятности продолжения $Q(t)$ для рассматриваемой грамматики стремятся к нулю при $t \to 0$ в силу согласованности грамматики, то $x(t), y(t) \to 0$ при $t \to \infty$. Следовательно, $E_t \to 0$ при $t \to \infty$ ввиду оценки $E_t = O(\max(x(t), y(t))$. Поскольку $Q(t) \ge 0$, то, применяя утверждение (2) леммы III.2.3, получаем, что при $t \to \infty$

$$x_i(t) = u_i' x_*(t) \cdot (1 + o(1)), \tag{IV.1.1.10}$$

и $y_i(t)/x_j(t) = o(1)$.

Вычитая из соотношения (IV.1.1.3) такое же соотношение для $t+1$, и пользуясь равенством

$$\delta\big(z_1(t) \cdot \ldots \cdot z_d(t)\big) = \sum_{i=1}^{d} z_1(t+1) \cdot \ldots \cdot z_{i-1}(t+1) \cdot \delta z_i(t) \cdot z_{i+1}(t) \cdot \ldots \cdot z_d(t), \tag{IV.1.1.11}$$

получаем соотношение

$$P(t+1) = -(\delta x(t), \delta y(t)) = (A - E'_t) \cdot P(t),$$

где $E'_t = O(x(t))$, и $E'_t \geq 0$ при достаточно большом t. Отсюда, пользуюсь тем, что $P(t) > 0$, получаем при $t \to \infty$ соотношение

$$\delta x(t) = u'_i \delta x_*(t) \cdot (1 + o(1)). \qquad \text{(IV.1.1.12)}$$

Из соотношений (IV.1.1.5) для $y_i(t), \delta y_i(t)$ следуют соотношения

$$\begin{aligned} y_i(t) &= u''_{i-k_1} y_*(t) \cdot (1 + o(1)), \\ \delta y_i(t) &= u''_{i-k_1} \delta y_*(t) \cdot (1 + o(1)). \end{aligned} \qquad \text{(IV.1.1.13)}$$

Лемма IV.1.1.3 *Для величины $x(t)$, заданной рекуррентными соотношениями (IV.1.1.3), (IV.1.1.4), где $x(t) \to 0, y(t)/x(t) \to 0$ при $t \to \infty$, выполнены асимптотическое равенства*

$$x_i(t) \sim u'_i k_0 t^{-1/2}, \; x_i(t) - x_i(t+1) \sim u'_i k_0 t^{-3/2}/2,$$

где $k_0 = \sqrt{\frac{4b}{B_1 B_2}}$, и величины B_1, B_2 заданы равенствами

$$\begin{aligned} B_1 &= \textstyle\sum_{i,j,l \leq k_1} v'_i b^i_{jl} u'_j u'_l, \\ B_2 &= \textstyle\sum_{i,j,l > k_1} v''_{i-k_1} b^i_{jl} u''_{j-k_1} u''_{l-k_1}, \end{aligned} \qquad \text{(IV.1.1.14)}$$

причем $B_1 B_2 > 0$, и $b = v' A^{(2)} u''$ введено ранее для докритического случая.

Сначала получим оценку $x_*(t) \sim k_0 t^{-1/2}$. Для этого покажем, что $t \delta x_*(t) \to 0$ при $t \to \infty$.

Домножим (скалярно) равенство (IV.1.1.3) слева на v', и произведем подстановки величин $x(t), y(t)$ из соотношений (IV.1.1.10), (IV.1.1.13). Пользуясь тем, что $v' A^{(1)} = v'$, получаем соотношение

$$\delta x_*(t) = v' A^{(2)} u'' \cdot y_*(t) \cdot (1 + o(1)) - \sum_{i,j,l \leq k_1} v'_i b^i_{jl} u'_j u'_l \cdot x_*^2(t)(1 + o(1))/2. \qquad \text{(IV.1.1.15)}$$

Подставим в равенство (IV.1.1.3) $t = t+1$ и вычтем из него аналогичное равенство для t. Домножая слева на v', и учитывая, что $v' A^{(1)} = v'$, получаем соотношение

$$\delta x_*(t+1) = \delta x_*(t) + v' A^{(2)} \delta y(t) \cdot (1 + O(x_*(t)))$$

$$- \sum_{i,j,l \leq k_1} v'_i b^i_{jl} \cdot \Big(x_j(t+1) \delta x_l(t) + x_l(t) \delta x_j(t) \Big) \cdot \Big(1 + O\big(x_*(t) + y_*(t)/x_*(t)\big) \Big)/2.$$

Подставляя выражения (IV.1.1.10),(IV.1.1.12) для $x_i(t), \delta x_i(t)$ и (IV.1.1.13) для $y_i(t), \delta y_i(t)$, и используя оценку $\delta x_*(t) = O(x_*^2(t) + y_*(t))$, вытекающую из равенства (IV.1.1.15), получаем

$$\delta x_*(t+1) = \delta x_*(t) + v' A^{(2)} u'' \delta y_*(t) \cdot \Big(1 + o(1) + O(x_*(t)) \Big)$$

$$-x_*(t)\delta x_*(t)\cdot\Big(1+o(1)+O\big(x_*(t)+y_*(t)/x_*(t)\big)\Big)ot\sum_{i,j,l\le k_1}v_i'b_{jl}^iu_j'u_l'. \qquad \text{(IV.1.1.16)}$$

Рассмотрим величину $z(t) = t\delta x_*(t)$. Подставляя $\delta x_*(t) = z(t)/t$, выражения (IV.1.1.10),(IV.1.1.12),(IV.1.1.13) для $x_i(t), \delta x_i(t), y_i(t), \delta y_i(t)$ в уравнение (IV.1.1.16), и учитывая, что $y_*(t)/x_*(t) \to 0$, $\delta x_*(t) = o(x_*(t))$, имеем

$$\frac{\delta z(t)}{t+1}-\frac{z(t)}{t(t+1)}=v'A^{(2)}u''\delta y_*(t)\cdot(1+o(1))-B_1x_*(t)z(t)\cdot(1+o(1))/t.$$

Домножая на $t+1$, учитывая оценку $\frac{1}{tx(t)} = o(1)$, вытекающую из соотношения $y_*(t)/x_*(t) \to 0$ и оценки (IV.1.1.5) для $y_*(t)$, и подставляя $\delta y_*(t) = (2+o(1))/(B_2t^2)$, получаем соотношение

$$\delta z(t) = -2b\cdot(1+o(1))/(B_2t)-B_1x_*(t)z(t)\cdot(1+o(1)). \qquad \text{(IV.1.1.17)}$$

Опять пользуясь оценкой $\frac{1}{tx(t)} = o(1)$, получаем, что к величине $z(t)$ можно применить лемму IV.1.1.1. Поэтому $z(t) = o(1)$ при $t \to \infty$. Подставляя $z(t) = t\delta x_*(t)$ в уравнение (IV.1.1.15), имеем:

$$by_*(t)-B_1x_*^2(t)/2 = o(1/t),$$

откуда следует, что $x_*^2(t) = 4b\cdot(1+o(1))/(B_1B_2t)$. Используя соотношения (IV.1.1.10), получаем требуемую асимптотику для $x(t)$. Подставляя найденную оценку для $x_*(t)$ в равенство (IV.1.1.17) и пользуясь леммой IV.1.1.2, получаем, что $z(t) = 2b\cdot(1+o(1))/(B_1B_2x_*(t)t)$, откуда вытекает оценка

$$\delta x_*(t) = -k_0\cdot(1+o(1))/(2t^{-3/2}).$$

Из соотношения (IV.1.1.12) находим:

$$x_i(t)-x_{i+1}(t) = -\delta x_i(t) = u_i'k_0t^{-3/2}(1+o(1)).$$

Лемма доказана.

как Из этой леммы сразу же следует

Теорема IV.1.1.2 *Для КС-грамматики с двумя классами нетерминалов и матрицей первых моментов, имеющей вид (I.2.1) при $r' = r'' = 1$, $B_1B_2 > 0$, где B_1, B_2 введены формулами (IV.1.1.14), при $t \to \infty$ верны следующие асимптотические равенства:*

$$Q_i(t) = u_i'k_0t^{-1/2}\cdot(1+o(1)),$$

$$P_i(t) = P(D_i^t) = \frac{u_i'k_0\cdot(1+o(1))}{2t^{3/2}},\ i=1,\dots,k_1,$$

где $k_0 = \sqrt{\frac{4b}{B_1B_2}}$, и b введено формулой (III.2.3).

IV.1.2 Математические ожидания числа применений правил в деревьях вывода

Рассмотрим случайные величины $q_{ij}^l(t,\tau), \bar{q}_{ij}^l(t,\tau)$ - соответственно число применений правила r_{ij} в деревьях из D_l^t и $D_l^{\leq t}$ на ярусе τ. Далее, пусть

$$S_{ij}^l(t) = q_{ij}^l(t,1) + q_{ij}^l(t,2) + \ldots + q_{ij}^l(t,t-1),$$

$$\bar{S}_{ij}^l(t) = \bar{q}_{ij}^l(t,1) + \bar{q}_{ij}^l(t,2) + \ldots + \bar{q}_{ij}^l(t,t-1)$$

- число применений правила r_{ij} в дереве вывода из D_l^t и $D_l^{\leq t}$ соответственно. Будем обозначать

$$S_{ij}(t) = S_{ij}^1(t),\ \bar{S}_{ij}(t) = \bar{S}_{ij}^1(t),$$

$$q_{ij}(t,\tau) = q_{ij}^1(t,\tau),\ \bar{q}_{ij}(t,\tau) = \bar{q}_{ij}^1(t,\tau).$$

Для краткости обозначим $M_{ij}^l(t) = M(S_{ij}^l(t)),\ \bar{M}_{ij}^l(t) = M(\bar{S}_{ij}^l(t))$.

Используя установленные в теореме IV.1.1.2 свойства деревьев вывода, можно построить рекуррентные соотношения для величин $M_{ij}^l(t),\ \bar{M}_{ij}^l(t)$ при $l, i = 1,\ldots,k,\ j = 1,\ldots,n_i$.

Для решения этих соотношений нам понадобится следующая лемма, в некотором смысле усиленный вариант леммы III.2.3:

Лемма IV.1.2.1 *Пусть $A(t) \to A$ при $t \to \infty$ - последовательность матриц размером $k \times k$, причем матрица $A > 0$ и ее перронов корень $r = 1$. Пусть $b(t) = bt^\alpha(1 + o(1))$ при $t \to \infty$ - последовательность векторов длины k, где $b \geq 0, b \neq 0$ - вектор длины k, а α - произвольное действительное число. Тогда для последовательности векторов $x(t), t = 0,1,\ldots,$ определяемой рекуррентным соотношением $x(t) = b(t) + A(t)x(t-1)$, при $t \to \infty$ справедливо соотношение*

$$\frac{x_i(t)}{vx(t)} \to u_i$$

при условии, что $x(t_0) > 0$ для некоторого номера t_0, где $v, u > 0$ - соответственно левый и правый собственные вектора матрицы A при нормировке $vu = 1$.

Для доказательства утверждения леммы достаточно показать, что для любого $\varepsilon > 0$ существует номер $t_1 > t_0$, такой, что $|x(t) - uvx(t)| < \varepsilon uvx(t)$ при $t > t_1$. Зафиксируем произвольно малые величины $0 < \varepsilon_1, \varepsilon_2 < 1$. Возьмем такое $t_2 = t_2(\varepsilon_1)$, что $|A(t) - A| < \varepsilon_1 A,\ |b(t) - bt^\alpha| < bt^\alpha\varepsilon_1$ при $t \geq t_2$. Поскольку $A^t \to uv$ при $t \to \infty$ для $A > 0$ [16], то существует $n_1 = n_1(\varepsilon_2)$, такое что $|A^t - uv| < uv\varepsilon_2$ при $t \geq n_1$. Положим $n_2 = n_2(\varepsilon_2) = n_1/\varepsilon_2$.

Введем обозначение $A^*(m,n) = \prod\limits_{i=0}^{m-n-1} A(m-i)$ (при $m > n$), причем считаем, что $A^*(n,n) = E$. Очевидно,

$$x(t+t_2) = A^*(t+t_2,t_2)x(t_2) + \sum_{m=1}^{t} A^*(t+t_2,m+t_2)b(m+t_2). \qquad \text{(IV.1.2.1)}$$

В силу неравенства $(1-\varepsilon_1)A < A(t) < (1+\varepsilon_1)A$, выполняющегося при $t \geq t_2$, получаем, что при $n \geq t_2$ справедливо неравенство

$$(1-\varepsilon_1)^m \cdot A^m < A^*(m+n,n) < (1+\varepsilon_1)^m \cdot A^m. \qquad \text{(IV.1.2.2)}$$

Положим в равенстве (IV.1.2.1) $t = n_2$. Обозначим

$$\begin{aligned} S_0 &= A^*(n_2+t_2,t_2)x(t_2),\\ S_1 &= \sum_{m=1}^{n_2-n_1} A^*(n_2+t_2,m+t_2)b(m+t_2),\\ S_2 &= \sum_{m=n_2-n_1+1}^{n_2} A^*(n_2+t_2,m+t_2)b(m+t_2). \end{aligned}$$

Вклад первого слагаемого S_0 в $x(n_2+t_2) - uvx(n_2+t_2)$ можно оценить как

$$|A^*(n_2+t_2,t_2)x(t_2) - uvA^*(n_2+t_2,t_2)x(t_2)| \leq |A^*(n_2+t_2,t_2)x(t_2) - A^{n_2}x(t_2)|$$

$$+|A^{n_2}x(t_2) - uvA^{n_2}x(t_2)| + |uvA^{n_2}x(t_2) - uvA^*(n_2+t_2,t_2)x(t_2)|.$$

Ввиду неравенства $|A^{n_2} - uv| < uv\varepsilon_2$ и неравенств (IV.1.2.2) можно записать

$$|A^{n_2}x(t_2) - uvA^{n_2}x(t_2)| < |(1+\varepsilon_2)uvx(t_2) - (1-\varepsilon_2)uvx(t_2)| < 2\varepsilon_2(1-\varepsilon_2)^{-1}A^{n_2}x(t_2)$$

$$\leq 2\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2}A^*(n_2+t_2,t_2)x(t_2).$$

Используя неравенство $uvx > \min(uv)x$, и обозначая $M_{uv} = (\min(uv))^{-1}$, имеем неравенство

$$|A^{n_2}x(t_2) - uvA^{n_2}x(t_2)| < 2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2}uvA^*(n_2+t_2,t_2)x(t_2).$$

Опять применяя неравенства (IV.1.2.2), получаем оценку

$$|A^*(n_2+t_2,t_2)x(t_2) - A^{n_2}x(t_2)| + |uvA^{n_2}x(t_2) - uvA^*(n_2+t_2,t_2)x(t_2)|$$

$$\begin{aligned} &< (1+M_{uv}) \cdot |uvA^*(n_2+t_2,t_2)x(t_2) - uvA^{n_2}x(t_2)|\\ &< (1+M_{uv})\left((1-\varepsilon_1)^{-n_2} - 1\right) uvA^*(n_2+t_2,t_2)x(t_2). \end{aligned}$$

Поэтому

$$|S_0 - uvS_0| < \Big(2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2} + (1+M_{uv}) \cdot \big((1-\varepsilon_1)^{-n_2} - 1\big)\Big) \cdot uvS_0. \qquad \text{(IV.1.2.3)}$$

Вклад в $x(n_2+t_2) - uvx(n_2+t_2)$ слагаемых из S_1 оценивается с помощью неравенств (IV.1.2.2) таким же образом как

$$|A^*(n_2+t_2,m+t_2)b(m+t_2) - uvA^*(n_2+t_2,m+t_2)b(m+t_2)| \leq$$

$$\Big(2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2} + (1+M_{uv}) \cdot \big((1-\varepsilon_1)^{-n_2} - 1\big)\Big) \cdot uvA^*(n_2+t_2,m+t_2)b(m+t_2),$$

откуда получаем оценку

$$|S_1 - uvS_1| < \Big(2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2} + (1+M_{uv}) \cdot \big((1-\varepsilon_1)^{-n_2} - 1\big)\Big) \cdot uvS_1. \qquad \text{(IV.1.2.4)}$$

Поскольку матрицы $A^m,\ m \geq 0$, положительны и ограничены, можно считать, что $M_1 < A^m < M_2$ для некоторых констант $M_1, M_2 > 0$ при любом $m \geq 0$. Оценим вклад в $x(n_2+t_2)-uvx(n_2+t_2)$ слагаемых из S_2. Докажем, что он мал по сравнению со всей суммой $S_1+S_2 = \sum\limits_{m=1}^{n_2} A^*(n_2+t_2, m+t_2)b(m+t_2)$. Опять применяя оценку (IV.1.2.2), и пользуясь оценкой $b(m+t_2) < (1+\varepsilon_1)b\cdot(m+t_2)^\alpha$, получим:

$$|S_2 - uvS_2| \leq (1+k\max(uv))\cdot\max(S_2)$$

$$\leq (1+k\max(uv))\cdot(1+\varepsilon_1)^{n_1}\max\left(\sum_{m=0}^{n_1-1}A^m b\right)\cdot\max_{m=n_2-n_1+1,\dots,n_2}(m+t_2)^\alpha$$

$$\leq \big(1+k\max(uv)\big)kM_2n_1\cdot(1+\varepsilon_1)^{n_1}\max(b)\cdot x_{m=n_2-n_1+1,\dots,n_2}(m+t_2)^\alpha. \qquad \text{(IV.1.2.5)}$$

Оценим сумму S_1+S_2 снизу:

$$S_1+S_2 = \sum_{m=1}^{n_2}A^*(n_2+t_2,m+t_2)b(m+t_2) > (1-\varepsilon_1)^{n_2}\cdot\min_{m=1,\dots,n_2}(m+t_2)^\alpha\cdot\sum_{m=0}^{n_2-1}A^m b$$

$$> (1-\varepsilon_1)^{n_2}n_2M_1b\cdot\min_{m=1,\dots,n_2}(m+t_2)^\alpha.$$

Поэтому

$$uv(S_1+S_2) > (1-\varepsilon_1)^{n_2}M_{uv}n_2M_1\max(b)\cdot\min_{m=1,\dots,n_2}(m+t_2)^\alpha.$$

в силу неравенства $uvx > \min(uv)\max(x)$. Умножая и деля правую часть оценки (IV.1.2.5) на правую часть полученного неравенства, получаем оценку

$$|S_2-uvS_2| \leq \frac{n_1kM_2(1+\varepsilon_1)^{n_1}\cdot(1+k\max(uv))\cdot\max\limits_{m=n_2-n_1+1,\dots,n_2}(m+t_2)^\alpha}{M_{uv}n_2M_1(1-\varepsilon_1)^{n_2}\cdot\min\limits_{m=1,\dots,n_2}(m+t_2)^\alpha} otuv(S_1+S_2)$$

Учитывая, что $n_1/n_2 = \varepsilon_2$, и применив очевидное неравенство

$$\frac{\max\limits_{m=n_2-n_1+1,\dots,n_2}(m+t_2)^\alpha}{\min\limits_{m=1,\dots,n_2}(m+t_2)^\alpha} \leq \Big(\frac{t_2+n_2}{t_2}\Big)^{|\alpha|},$$

имеем оценку

$$|S_2-uvS_2| \leq \frac{\varepsilon_2kM_2\cdot(1+k\max(uv))}{M_{uv}M_1}\cdot\Big((1+\varepsilon_1)^{n_1}/(1-\varepsilon_1)^{n_2}\Big)$$

$$\cdot\Big(\frac{t_2+n_2}{t_2}\Big)^{|\alpha|}uv(S_1+S_2). \qquad \text{(IV.1.2.6)}$$

Суммируя неравенства (IV.1.2.3), (IV.1.2.4), (IV.1.2.6), и пользуясь неравенством $|A+B| < |A|+|B|$, получим оценку

$$|x(n_2+t_2)-uvx(n_2+t_2)| \leq |S_0-uvS_0|+|S_1-uvS_1|+|S_2-uvS_2| <$$

$$\Big(2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2}+(1+M_{uv})\cdot\big((1-\varepsilon_1)^{-n_2}-1\big)\Big)\cdot uvS_0+$$
$$\Big[\frac{\varepsilon_2 kM_2\cdot(1+k\max(uv))}{M_{uv}M_1}\cdot\Big((1+\varepsilon_1)^{n_1}/(1-\varepsilon_1)^{n_2}\Big)\cdot\Big(\frac{t_2+n_2}{t_2}\Big)^{|\alpha|}+$$
$$\Big(2M_{uv}\varepsilon_2(1-\varepsilon_2)^{-1}(1-\varepsilon_1)^{-n_2}+(1+M_{uv})\cdot\big((1-\varepsilon_1)^{-n_2}-1\big)\Big)\Big]\cdot uv(S_1+S_2).$$

Учитывая, что $x(n_2+t_2)=S_0+S_1+S_2$, получим для некоторой константы $C>0$, не зависящей от t_2, n_2:

$$|x(n_2+t_2)-uvx(n_2+t_2)|<C\cdot\Big(\big((1-\varepsilon_1)^{-n_2}-1\big)+2\varepsilon_2(1-\varepsilon_2)^{-1}\cdot\Big(\frac{1+\varepsilon_1}{1-\varepsilon_1}\Big)^{n_2}$$
$$\times\Big(\frac{t_2+n_2}{t_2}\Big)^{|\alpha|}\Big)uvx(n_2+t_2).$$

Устремляя сначала $\varepsilon_2\to 0$, а затем $\varepsilon_1\to 0$ и $t_2=t_2(\varepsilon_1)\to\infty$, получим соответствующие номера t_2, n_2, такие что $|x(n_2+t)-uvx(n_2+t)|<\varepsilon uvx(n_2+t)$ для сколь угодно малого ε при $t>t_2$. Лемма доказана.

Рассмотрим величину $\bar{M}^q_{ij}(t)$. Пусть $D^{\leq t}_{ql}$ - множество деревьев из $D^{\leq t}_q$, корень которых помечен нетерминалом A_q, и в нем применено правило r_{ql}. Обозначим через $\bar{P}^{ij}_{ql}(t)$ величину $\sum\limits_{d\in D^{\leq t}_{ql}} p(d)q_{ij}(d)$, где $p(d)$ - условная вероятность дерева d в $D^{\leq t}_q$, а $q_{ij}(d)$ - общее число применений правила r_{ij} в дереве d. Очевидно, $\bar{M}^q_{ij}(t)=\sum_l \bar{P}^{ij}_{ql}(t)$. В дальнейшем в обозначении $\bar{P}^{ij}_{ql}(t)$ будем для краткости опускать индексы i,j.

Напомним, что вероятность наборов деревьев вывода $D^{\leq t}_X$ (где $X=(x_1,\ldots,x_k)$), в которых ровно x_i деревьев имеют корень, помеченный нетерминалом A_i для $i=1,\ldots,k$, и высоты которых не превосходят t, обозначается $Q_X(t)$, а вероятность наборов деревьев D^t_X из $D^{\leq t}_X$, в которых хотя бы одно достигает яруса t, обозначается $R_X(t)$. Обозначим за $s_{ij}=(s^1_{ij},\ldots,s^k_{ij})$ вектор нетерминалов в правой части правила r_{ij}.

Рассмотрим величину $\bar{P}_{ql}(t)$. Можно записать $q_{ij}(d)=q^{(1)}_{ij}(d)+q^{(2)}_{ij}(d)$, где $q^{(1)}_{ij}(d)$ равно единице, если $r_{ij}=r_{ql}$, и нулю в противном случае, а $q^{(2)}_{ij}(d)$ обозначает количество применений правила r_{ij} в поддеревьях с корнем на первом ярусе. Очевидно, что

$$\bar{P}_{ql}(t)=\sum_{d\in D^{\leq t}_{ql}} p(d)q_{ij}(d)=\sum_{d\in D^{\leq t}_{ql}} p(d)q^{(1)}_{ij}(d)+\sum_{d\in D^{\leq t}_{ql}} p(d)q^{(2)}_{ij}(d)=\bar{P}^1_{ql}(t)+\bar{P}^2_{ql}(t).$$

Таким образом, $\bar{P}^1_{ql}(t)$ учитывает только применение правила r_{ij} в корне дерева, а $\bar{P}^2_{ql}(t)$ учитывает количество применений правила r_{ij} к остальным вершинам дерева. Очевидно, что $\bar{P}^1_{ql}(t)=p_{ij}Q_{s_{ij}}(t-1)/(1-Q_q(t))$ при $i=q, j=l$, и $\bar{P}^1_{ql}(t)=0$ в противном случае.

Обозначим через $\delta^i(n)=(\delta^i_1,\ldots,\delta^i_n)$ вектор длины n, в котором на позиции i стоит единица, а все остальные компоненты равны нулю. Параметр n будет опускаться, если из контекста длина вектора известна.

Вероятность дерева $d \in D_{ql}^{\leq t}$ можно представить в виде $p(d) = p_{ql}p_1(d)p_2(d) \cdot \ldots \cdot p_{\bar{s}_{ql}}(d)$, где $p_i(d)$ - вероятность поддерева d с i-м по порядку корнем на втором ярусе. Количество применений правила r_{ij} в дереве d на ярусах с номерами, большими единицы $q'_{ij}(d)$ можно представить в виде суммы $q'_{ij}(d) = \sum_{m=1}^{\bar{s}_{ql}} q'^{m}_{ij}(d)$, где $q'^{m}_{ij}(d)$ - количество применений правила r_{ij} в поддереве с m-м по порядку корнем на втором ярусе.

Очевидно,

$$\bar{P}_{ql}^2(t) = \frac{1}{1-Q_q(t)} \cdot \sum_d p(d)q'_{ij}(d) = \frac{p_{ql}}{1-Q_q(t)} \cdot \sum_d \prod_{n=1}^{\bar{s}_{ql}} p_n(d) \sum_{m=1}^{\bar{s}_{ql}} q'^{m}_{ij}(d)$$

$$= \frac{p_{ql}}{1-Q_q(t)} \cdot \sum_{m=1}^{\bar{s}_{ql}} \sum_d p_1(d)\ldots p_{m-1}(d)p_{m+1}(d)\ldots p_{\bar{s}_{ql}}(d) \cdot p_m(d)\bar{M}_{ij}^m(t-1)$$

$$= \frac{p_{ql}}{1-Q_q(t)} \cdot \sum_m s_{ql}^m \bar{M}_{ij}^m(t-1) \cdot Q_{s_{ql}}(t-1).$$

Суммируя по l, получаем следующее рекуррентное соотношение для $\bar{M}_{ij}^q(t)$:

$$\bar{M}_{ij}^q(t) = \frac{1}{1-Q_q(t)} \cdot \left(\delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_l p_{ql} \sum_m s_{ql}^m \bar{M}_{ij}^m(t-1) Q_{s_{ql}}(t-1)\right). \tag{IV.1.2.7}$$

Обозначим $\bar{M}'^q_{ij}(t) = \bar{M}_{ij}^q(t) \cdot (1 - Q_q(t))$. Тогда соотношение (IV.1.2.7) можно записать в виде

$$\bar{M}'^q_{ij}(t) = \delta_i^q p_{ij} Q_{s_{ij}}(t-1) + \sum_l p_{ql} \sum_m s_{ql}^m \bar{M}'^m_{ij}(t-1) Q_{s_{ql}-\delta^{z_m}}(t-1). \tag{IV.1.2.8}$$

Отметим, что $\bar{M}'^q_{ij}(t) = \bar{M}_{ij}^q(t) \cdot (1 + o(1))$.

Аналогично соотношению для $\bar{M}_{ij}^q(t)$ получаем рекуррентное соотношение для $M_{ij}^q(t)$. Вклад $P_{ql}^1(t)$ правила, примененного к корню дерева, равен $\delta_i^q p_{ij} R_{s_{ij}}(t-1)/P_q(t)$. Вклад $P_{ql}^2(t)$ деревьев из $D_{s_{ql}}^t$ можно представить в виде суммы $\sum_m P_{ql}^{2m}(t)$, где $P_{ql}^{2m}(t)$ учитывает вклад дерева, корень которого помечен m-м по порядку нетерминалом $z_m, m = 1, \ldots, \bar{s}_{ql}$, стоящим в правой части правила r_{ql}. Вклад в $P_{ql}^{2m}(t)$ наборов деревьев из $D_{s_{ql}}^{t-1}$, в которых ярус t достигается через дерево с корнем на первом ярусе, помеченным нетерминалом z_m, равен

$$S_1 = P_{z_m}(t-1) Q_{s_{ql}-\delta^{z_m}}(t-1) M_{ij}^{z_m}(t-1)/P_q(t),$$

а вклад наборов, в которых ярус t достигается через другие деревья, равен

$$S_2 = (1 - Q_{z_m}(t-1)) R_{s_{ql}-\delta^{z_m}}(t-1) \bar{M}_{ij}^{z_m}(t-1)/P_q(t).$$

Подставляя $P_{ql}^{2m}(t) = S_1 + S_2$, получаем равенство

$$M_{ij}^q(t) = \sum_{l=1}^{n_q} \left(P_{ql}^1(t) + \sum_{m=1}^{\bar{s}_{ql}} P_{ql}^{2m}(t)\right)$$

$$= \frac{1}{P_q(t)} \cdot \Big(\delta_i^q p_{ij} R_{s_{ij}}(t-1) + \sum_l p_{ql} \sum_m s_{ql}^m \cdot \big(P_m(t-1) Q_{s_{ql}-\delta^m}(t-1) M_{ij}^m(t-1)$$

$$+(1-Q_m(t-1)) R_{s_{ql}-\delta^m}(t-1) \bar{M}_{ij}^m(t-1)\big)\Big). \qquad \text{(IV.1.2.9)}$$

Из леммы III.4.1 вытекают равенства

$$\begin{aligned} Q_X(t) &= \prod_{i=1}^k (1-Q_i(t))^{x_i} = 1 - \sum_{i=1}^k x_i Q_i(t) + O\left(\textstyle\sum_{i,j} x_i x_j Q_i(t) Q_j(t)\right), \\ R_X(t) &= Q_X(t) - Q_X(t-1) = \sum_{i=1}^k x_i P_i(t) + O\left(\textstyle\sum_{i,j} x_i x_j Q_i(t) P_j(t)\right), \end{aligned} \qquad \text{(IV.1.2.10)}$$

применяя которые к выражениям для $Q_i(t), P_i(t)$, полученным в теореме IV.1.1.2, получаем оценки

$$Q_{s_{ij}-\delta^m}(t) = \begin{cases} 1 - \sum_{l \le k_1} (s_{ij}^l - \delta_l^m) Q_l(t) + o(t^{-1/2}), & i \le k_1, \\ 1 - \sum_{l > k_1} (s_{ij}^l - \delta_l^m) Q_l(t) + o(1/t), & i > k_1, \end{cases} \qquad \text{(IV.1.2.11)}$$

$$R_{s_{ij}-\delta^m}(t) = \begin{cases} \sum_{l \le k_1} (s_{ij}^l - \delta_l^m) P_l(t) + o(t^{-3/2}), & i \le k_1, \\ \sum_{l > k_1} (s_{ij}^l - \delta_l^m) P_l(t) + o(1/t^2), & i > k_1. \end{cases} \qquad \text{(IV.1.2.12)}$$

Найдем сначала $\bar{M}_{ij}^q(t), \bar{M}_{ij}^{\prime q}(t)$ при $i, q > k_1$. Подставляя оценки (IV.1.2.11), (IV.1.2.12) и полученные ранее выражения для $Q_i(t)$ в соотношение (IV.1.2.8), и учитывая, что $(1-Q_i(t))^{-1} = 1 + Q_i(t) + O(Q_i^2(t))$, получаем следующее соотношение для величин $\bar{M}_{ij}^{\prime q}(t)$:

$$\bar{M}_{ij}^{\prime q}(t) = \delta_i^q p_{ij} + \sum_{m > k_1} \sum_l p_{ql} s_{ql}^m M_{ij}^{\prime m}(t-1)$$

$$- \sum_{m,n > k_1} \sum_l p_{ql} s_{ql}^m (s_{ql}^n - \delta_n^m) \cdot Q_n(t-1) \bar{M}_{ij}^{\prime m}(t-1) \cdot (1 + o(1))$$

$$= \delta_i^q p_{ij} + \sum_{m > k_1} a_m^q \bar{M}_{ij}^{\prime m}(t-1) - \frac{2 + o(1)}{B_2 t} \cdot \sum_{m,n > k_1} b_{mn}^q u_{n-k_1}'' \bar{M}_{ij}^{\prime m}(t-1). \qquad \text{(IV.1.2.13)}$$

В последнем преобразовании были использованы равенства

$$a_m^q = \sum_l p_{ql} s_{ql}^m, \;\; b_{mn}^q = \sum_l p_{ql} s_{ql}^m (s_{ql}^n - \delta_n^m).$$

Применяя лемму IV.1.2.1, получаем, что

$$\bar{M}_{ij}^{\prime q}(t) = u_{q-k_1}'' \cdot (1 + o(1)) \cdot \sum_{n > k_1} v_{n-k_1}'' \bar{M}_{ij}^n(t). \qquad \text{(IV.1.2.14)}$$

Если обозначить $\bar{M}_{ij}^*(t) = \sum_{n > k_1} v_{n-k_1}'' \bar{M}_{ij}^{\prime n}(t)$, то, домножив соотношение (IV.1.2.13) слева на v_{q-k_1}'', суммируя по $q > k_1$ и используя равенства (IV.1.2.14), получим:

$$\delta \bar{M}_{ij}^*(t) = v_{i-k_1}'' p_{ij} - \frac{(2 + o(1)) \bar{M}_{ij}^*(t)}{B_2 t} \cdot \sum_{m,n,q > k_1} v_{q-k_1}'' b_{mn}^q u_{m-k_1}'' u_{n-k_1}''$$

$$= v''_{i-k_1} p_{ij} - \bar{M}^*_{ij}(t) \cdot (2 + o(1))/t.$$

Применяя лемму IV.1.1.2, получаем $\bar{M}^*_{ij}(t) = v''_{i-k_1} p_{ij} t \cdot (1 + o(1))/3$, то есть

$$\begin{aligned} \bar{M}'^q_{ij}(t) &= u''_{q-k_1} v''_{i-k_1} p_{ij} t \cdot (1 + o(1))/3 \\ \bar{M}^q_{ij}(t) &= u''_{q-k_1} v''_{i-k_1} p_{ij} t \cdot (1 + o(1))/3. \end{aligned} \qquad \text{(IV.1.2.15)}$$

Теперь найдем $M^q_{ij}(t)$ при $i, q > k_1$. Обозначим далее $M'^q_{ij}(t) = M^q_{ij}(t) P_q(t)$. Из равенств (IV.1.2.9) с использованием оценок (IV.1.2.11), (IV.1.2.12), выражений для $Q_i(t), P_i(t)$, и доказанного равенства (IV.1.2.15) для $\bar{M}^m_{ij}(t)$, получаем соотношение

$$M'^q_{ij}(t) = O(p_{ij} t^{-3/2}) + \sum_{m>k_1} \sum_l p_{ql} s^m_{ql} M'^m_{ij}(t-1)$$

$$- \sum_{m,n>k_1} \sum_l p_{ql} s^m_{ql} (s^n_{ql} - \delta^m_n) \cdot Q_n(t-1) M'^m_{ij}(t-1)$$

$$+ \sum_{m,n>k_1} \sum_l p_{ql} s^m_{ql} (s^n_{ql} - \delta^m_n) \cdot P_n(t-1) \bar{M}^m_{ij}(t-1)$$

$$= \sum_{m>k_1} a^q_m M'^m_{ij}(t-1) - \frac{2 + o(1)}{B_2 t} \cdot \sum_{m,n>k_1} b^q_{mn} u''_{n-k_1} M'^m_{ij}(t-1)$$

$$+ \frac{p_{ij} v''_{i-k_1} \cdot (2 + o(1))}{3 B_2 t} \cdot \sum_{m,n>k_1} b^q_{mn} u''_{n-k_1} u''_{m-k_1}.$$

Применяя лемму IV.1.2.1, имеем:

$$M'^q_{ij}(t) = u''_{q-k_1} \cdot \sum_{n>k_1} v''_{n-k_1} M'^n_{ij}(t) \cdot (1 + o(1)).$$

Домножая полученное рекуррентное соотношение для $M'^q_{ij}(t)$ на v''_{q-k_1}, суммируя по $q > k_1$ и обозначая $M^*_{ij}(t) = \sum_{n>k_1} v''_{n-k_1} M'^n_{ij}(t)$, аналогично предыдущему случаю получаем рекуррентное соотношение

$$\delta M^*_{ij}(t) = (2 + o(1)) p_{ij} v''_{i-k_1}/(3t) - (2 + o(1)) M^*_{ij}(t)/t,$$

откуда получаем $M^*_{ij}(t) = p_{ij} v''_{i-k_1} \cdot (1 + o(1))/3$, и

$$M^q_{ij}(t) = \frac{p_{ij}}{3 P_q(t)} u''_{q-k_1} v''_{i-k_1} \cdot (1 + o(1)) = p_{ij} v''_{i-k_1} B_2 t^2 \cdot (1 + o(1))/6.$$

Эта формула совпадает с соответствующей формулой, выведенной в [10] для неразложимой грамматики.

Аналогично для $\bar{M}'^q_{ij}(t)$ при $i, q \le k_1$ получаем соотношение

$$\bar{M}'^q_{ij}(t) = \delta^q_i p_{ij} + \sum_{m \le k_1} a^q_m \bar{M}'^m_{ij}(t-1)$$

$$-k_0(t-1)^{-1/2}(1+o(1))\cdot\sum_{m,n\le k_1} b^q_{mn}u'_n\bar M'^m_{ij}(t-1).$$

Применяя лемму IV.1.2.1, имеем:

$$\bar M'^q_{ij}(t)=u'_q\cdot\sum_{n\le k_1} v'_n\bar M'^n_{ij}(t)\cdot(1+o(1)).$$

Домножая рекуррентное соотношение для $\bar M'^q_{ij}(t)$ на v' и обозначая $\bar M^*_{ij}(t)=\sum_{n\le k_1} v'_n\bar M'^n_{ij}(t)$, получаем для $\bar M^*_{ij}(t)$ рекуррентное соотношение

$$\delta\bar M^*_{ij}(t)=v'_ip_{ij}-B_1k_0\bar M^*_{ij}(t)\cdot(1+o(1))t^{-1/2},$$

откуда находим $\bar M^*_{ij}(t)=v'_ip_{ij}t^{1/2}\cdot(1+o(1))/(B_1k_0)$, и

$$\begin{aligned}M'^q_{ij}(t)&=u'_qv'_ip_{ij}\cdot t^{1/2}(1+o(1))/(B_1k_0),\\ M^q_{ij}(t)&=u'_qv'_ip_{ij}\cdot t^{1/2}(1+o(1))/(B_1k_0).\end{aligned}$$

Рассмотрим величины $M^q_{ij}(t),\ i,q\le k_1$. Рекуррентное соотношение для $M'^q_{ij}(t)$ имеет вид:

$$M'^q_{ij}(t)=O(p_{ij}t^{-3/2})+\sum_{m\le k_1} a^q_mM'^m_{ij}(t-1)-k_0t^{-1/2}(1+o(1))\cdot\sum_{m,n\le k_1} b^q_{mn}u'_nM'^m_{ij}(t-1)$$

$$+\sum_{m,n\le k_1} b^q_{mn}P_n(t-1)\bar M^m_{ij}(t-1)(1+o(1))=$$

$$=\sum_{m\le k_1} a^q_mM'^m_{ij}(t-1)-k_0t^{-1/2}(1+o(1))\cdot\sum_{m,n\le k_1} b^q_{mn}u'_nM'^m_{ij}(t-1)$$

$$+\frac{v'_ip_{ij}\cdot(1+o(1))}{2tB_1}\cdot\sum_{m,n>k_1} b^q_{mn}u'_nu'_m.$$

По лемме IV.1.2.1,

$$M'^q_{ij}(t)=u'_q\cdot\sum_{n\le k_1} v'_nM'^n_{ij}(t)\cdot(1+o(1)).$$

Обозначая $M^*_{ij}(t)=\sum_{n\le k_1} v'_nM'^n_{ij}(t)$, получаем соотношение

$$\delta M^*_{ij}(t)=p_{ij}v'_i\cdot(1+o(1))/(2t)-B_1k_0t^{-1/2}(1+o(1))\cdot M^*_{ij}(t),$$

откуда следует асимптотическое равенство

$$M^q_{ij}(t)=p_{ij}u'_qv'_i\cdot(1+o(1))/(2B_1k_0t^{1/2}\cdot P_q(t))=v'_ip_{ij}t\cdot(1+o(1))/(B_1k_0^2).$$

Для $\bar M'^q_{ij}(t)$ при $q\le k_1,\ i>k_1$ имеем:

$$\bar M'^q_{ij}(t)=O(p_{ij})+\sum_{m\le k_1}\sum_l p_{ql}s^m_{ql}\bar M'^m_{ij}(t-1)\left(1-\sum_{n\le k_1}(s^n_{ql}-\delta^m_n)Q_n(t-1)\right)\cdot(1+o(1))$$

$$+\sum_{m>k_1}\sum_{l} p_{ql}s_{ql}^m \bar{M}_{ij}^{\prime m}(t-1)\cdot(1+o(1)).$$

Подставляя полученные выше выражения для $\bar{M}_{ij}^{\prime m}(t)$, $m,i>k_1$, и производя аналогичные предыдущим случаям преобразования, получаем равенство:

$$\bar{M}_{ij}^{\prime q}(t)=\sum_{m\le k_1} a_m^q \bar{M}_{ij}^{\prime m}(t-1)-k_0t^{-1/2}(1+o(1))\cdot\sum_{m,n\le k_1} b_{mn}^q u_n'\bar{M}_{ij}^{\prime m}(t-1)$$

$$+\frac{p_{ij}v_{i-k_1}''t\cdot(1+o(1))}{3}\cdot\sum_{m>k_1} a_m^q u_{m-k_1}''.$$

Отсюда, домножая на v_q' и суммируя, для $\bar{M}_{ij}^*(t)=\sum_{n\le k_1} v_n'\bar{M}_{ij}^{\prime m}(t)$ получаем соотношение

$$\delta\bar{M}_{ij}^*(t)=p_{ij}v_{i-k_1}''bt\cdot(1+o(1))/3-B_1k_0t^{-1/2}\bar{M}_{ij}^*(t)\cdot(1+o(1)),$$

(напомним, что $b=\sum_{q\le k_1,m>k_1} a_m^q v_q' u_{m-k_1}''$), откуда следует,что

$$\bar{M}_{ij}^{\prime q}(t)=p_{ij}bu_q'v_{i-k_1}''(1+o(1))t^{3/2}/(3B_1k_0),$$
$$\bar{M}_{ij}^{q}(t)=p_{ij}bu_q'v_{i-k_1}''(1+o(1))t^{3/2}/(3B_1k_0).$$

И, наконец, найдем величины $M_{ij}^q(t)$, $q\le k_1, i>k_1$. Рекуррентное соотношение для $M_{ij}^{\prime q}(t)$ имеет вид:

$$M_{ij}^{\prime q}(t)=O(t^{-3/2})+\sum_m a_m^q M_{ij}^{\prime m}(t-1)-\sum_{m<k,n\le k_1} b_{mn}^q Q_n(t)M_{ij}^{\prime m}(t-1)\cdot(1+o(1))$$

$$+\left(\sum_{m<k,n\le k_1} b_{mn}^q P_n(t-1)\bar{M}_{ij}^m(t-1)\right)\cdot(1+o(1)).$$

Заметим, что

$$\sum_{m>k_1,n\le k_1} b_{mn}^q Q_n(t)M_{ij}^{\prime m}(t-1)+\sum_{m>k_1,n\le k_1} b_{mn}^q P_n(t-1)\bar{M}_{ij}^m(t-1)=O(t^{-1/2}$$

в силу полученных ранее оценок для $\bar{M}_{ij}^m(t-1), M_{ij}^m(t-1)$ при $m,i>k_1$. Поэтому, подставляя $M_{ij}^{\prime m}(t-1)=P_m(t-1)M_{ij}^m(t-1)$ для $m>k_1$, получаем соотношение

$$M_{ij}^{\prime q}(t)=O(t^{-1/2})+\sum_{m\le k_1} a_m^q M_{ij}^{\prime m}(t-1)-\sum_{m,n\le k_1} b_{mn}^q Q_n(t)M_{ij}^{\prime m}(t-1)\cdot(1+o(1))$$

$$+\left(\sum_{m,n\le k_1} b_{mn}^q P_n(t-1)\bar{M}_{ij}^m(t-1)+\sum_{m>k_1} a_m^q P_m(t-1)M_{ij}^m(t-1)\right)\cdot(1+o(1)).$$

Используя найденные ранее выражения для $\bar{M}_{ij}^m(t)$, $m,i\le k_1$ и $M_{ij}^m(t)$, $m>k_1, i\le k_1$, и применяя лемму IV.1.2.1, получаем для $M_{ij}^*(t)=\sum_{n\le k_1} v_n' M_{ij}^{\prime m}(t)$ следующее рекуррентное соотношение:

$$\delta M_{ij}^*(t)=O(t^{-1/2})+\frac{bp_{ij}v_{i-k_1}''t^{3/2}}{3B_1k_0}\cdot\frac{k_0}{2t^{3/2}}\cdot\sum_{q,m\le k_1}\sum_{n\le k_1} v_q' b_{mn}^q u_n' u_m' + v_{i-k_1}''\cdot\sum_{m>k_1,q\le k_1} v_q' a_m^q u_{m-k_1}''$$

$$\cdot\frac{2}{B_2t^2}\cdot p_{ij}B_2t^2\cdot(1+o(1))/6-M^*_{ij}(t)B_1k_0t^{-1/2}\cdot(1+o(1))$$
$$=bp_{ij}v''_{i-k_1}\cdot(1+o(1))/2-M^*_{ij}(t)B_1k_0t^{-1/2}\cdot(1+o(1)),$$

откуда получаем

$$M^*_{ij}(t)=v''_{i-k_1}bp_{ij}t^{1/2}\cdot(1+o(1))/(2B_1k_0),$$
$$M^q_{ij}(t)=v''_{i-k_1}bp_{ij}t^2\cdot(1+o(1))/(B_1k_0^2).$$

Полученные результаты сформулируем в виде следующей теоремы.

Теорема IV.1.2.1 *Для разложимой грамматики с двумя классами нетерминалов, матрица первых моментов которой имеет вид (I.2.1) и $r'=r''=1$, $B_1B_2>0$, где B_1,B_2 введены формулами (IV.1.1.14), величины $M(S_{ij}(t))$ имеют следующий вид при $t\to\infty$:*

$$M(S_{ij}(t))=\frac{p_{ij}v'_iB_2t\cdot(1+o(1))}{4b},\ i\le k_1$$
$$M(S_{ij}(t))=p_{ij}v''_{i-k_1}B_2t^2\cdot(1+o(1))/4,\ i>k_1,$$

где b дано формулой (III.2.3), а p_{ij} - вероятность применения правила r_{ij} в грамматике.

Кратко суммируем полученные для случае кратного перронова корня $r=1$ свойства деревьев вывода t при $t\to\infty$.

1) Количество правил, примененных к нетерминалам второго класса, имеет такую же ($O(t^2)$) асимптотику, как и в неразложимом случае, а правил, примененных к первому нетерминалу, значительно меньше ($O(t)$).

2) Вероятность продолжения $Q_i(t)$ для $i\le k_1$ больше, чем $Q_i(t)$ для $i>k_1$.

3) Как и в неразложимом критическом случае, математическое ожидание количества применений правила r_{ij} в деревьях вывода пропорционально его вероятности p_{ij}, то есть, если за $M^i_j(t)=\sum_l M^i_{jl}(t)$ обозначить число нетерминалов A_j в дереве вывода из D^t_i, то $M^i_j(t)/M^i_{jl}(t)\to p_{jl}$ при $t\to\infty$.

4) Величины $M(S_{ij}(t))$ не меняются при замене аксиомы грамматики на любой нетерминал из класса K_1 .

IV.1.3 Энтропия и стоимость оптимального кодирования

В этом параграфе мы будем рассматривать только КС-грамматики с однозначным выводом. Сначала выведем асимптотическую формулу для энтропии множества слов $\mathcal{L}^t$ грамматики, имеющих деревья вывода высоты t. По определению, $H(\mathcal{L}^t)=-\sum\limits_{\alpha\in L^t}p_t(\alpha)\cdot\log p_t(\alpha)$. Как и в докритическом случае, мы можем записать, что

$$H(\mathcal{L}^t)=\log P(L^t)-\sum_{i=1}^{k}\sum_{j=1}^{n_i}\log p_{ij}\cdot M(S_{ij}(t)).$$

Во всех случаях, в силу доказанных выше оценок для $P(L^t) = P_1(t)$, имеем $\log P(L^t) = O(\log t)$. Пользуясь асимптотиками для $M(S_{ij}(t))$, полученными в теореме IV.1.2.1, получаем:

$$H(\mathcal{L}^t) = -\sum_{i\le k_1}\sum_{j=1}^{n_j}\log p_{ij}\cdot M(S_{ij}(t)) - \sum_{i>k_1}\sum_{j=1}^{n_j}\log p_{ij}\cdot M(S_{ij}(t)) + O(\log t)$$

$$= -\sum_{i>k_1} v''_{i-k_1}\cdot\sum_{j=1}^{n_j} p_{ij}\log p_{ij}\cdot B_2 t^2\cdot(1+o(1))/4 + O(t)$$

$$= \sum_{i>k_1} v''_{i-k_1} H(R_i)\cdot B_2 t^2\cdot(1+o(1))/4,$$

где

$$H(R_i) = -\sum_{j=1}^{n_j} p_{ij}\log p_{ij} \qquad \text{(IV.1.3.1)}$$

- энтропия множества правил R_i.

Таким образом, справедлива следующая теорема:

Теорема IV.1.3.1 *Для КС-грамматики с матрицей первых моментов вида (I.2.1) с $r = r' = r'' = 1$, $B_1B_2 > 0$, где B_1, B_2 введены формулами (IV.1.1.14), энтропия множества слов $\mathcal{L}^t$ имеет вид*

$$H(\mathcal{L}^t) = \sum_{i>k_1} v''_{i-k_1} H(R_i)\cdot B_2 t^2\cdot(1+o(1))/4,$$

где $H(R_i)$ определено формулой (IV.1.3.1).

Теперь найдем стоимость оптимального кодирования для случая кратного перронова корня $r = 1$. Пусть f^* - кодирование множества слов $\mathcal{L}^t$, минимизирующее величину $M_t(f)$, т.е. $M_t(f) \ge M_t(f^*)$ для любого кодирования f, определенного и однозначного на множестве слов $\mathcal{L}^t$. Как и в докритическом случае, поскольку $H(\mathcal{L}^t) \to \infty$ при $t \to \infty$, то из теоремы II.3 следует, что $M_t(f^*)/H(\mathcal{L}^t) \to 1$ при $t \to \infty$.

Рассмотрим величину $\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha|$. Пусть l_{ij} - количество терминальных символов в правой части правила r_{ij}. Ранее было показано, что

$$\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha| = \sum_{i=1}^{k}\sum_{j=1}^{n_i} l_{ij} M(S_{ij}(t)).$$

Подставляя полученные в теореме IV.1.2.1 асимптотики для $M(S_{ij}(t))$, находим, что

$$\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha| = \sum_{i>k_1} v''_{i-k_1} L(R_i)\cdot B_2 t^2(1+o(1))/4, \qquad \text{(IV.1.3.2)}$$

где

$$L(R_i) = \sum_{j=1}^{n_j} l_{ij} p_{ij} \qquad \text{(IV.1.3.3)}$$

- среднее число терминалов в правой части правил из R_i.

Как следствие теоремы IV.1.3.1, замечания о том, что $M_t(f^*)/H(\mathcal{L}^t) \to 1$ при $t \to \infty$ и асимптотического равенства (IV.1.3.2) верна следующая теорема:

Теорема IV.1.3.2 *Пусть $\mathcal{L}$ - КС-язык, порожденный разложимой грамматикой с матрицей первых моментов вида (I.2.1) при $r = r' = r'' = 1$, $B_1 B_2 > 0$. Тогда стоимость любого кодирования f этого языка удовлетворяет неравенству*

$$C(\mathcal{L}, f) \geq C^*(\mathcal{L}),$$

где

$$C^*(\mathcal{L}) = \frac{\sum_{i>k_1} v''_{i-k_1} H(R_i)}{\sum_{i>k_1} v''_{i-k_1} L(R_i)},$$

величины $H(R_i), L(R_i)$ введены формулами (IV.1.3.1),(IV.1.3.3), а B_1, B_2 - формулами (IV.1.1.14).

Заметим, что в случае кратного перронова корня нижняя оценка стоимости оптимального кодирования определяется свойствами правил, применяемых к нетерминалам из класса K_2, поскольку в этом случае математическое ожидание количества нетерминалов из K_2 в дереве вывода имеет больший порядок, чем математическое ожидание количества нетерминалов из K_1.

IV.1.4 Алгоритм оптимального кодирования

Можно показать, что полученная в предыдущем параграфе оценка для стоимости кодирования является точной, т. е. существует кодирование множества слов языка, стоимость которого равна $C^*(\mathcal{L})$. Такое кодирование строится следующим образом. Упорядочим слова языка в порядке невозрастания вероятностей, так что $p_1 \geq p_2 \geq \ldots$, где p_i-вероятность i-го слова. Поскольку

$$\sum_i 2^{-\lceil -\log p_i \rceil} \leq \sum_i 2^{\log p_i} = 1,$$

то из выполнения неравенства Мак-Миллана [22] следует, что существует взаимно-однозначное кодирование f_s множества слов языка, при котором длина кода i-го слова равна $\lceil -\log p_i \rceil$. Величина $M_t(f_s)$ для такого кодирования оценивается как

$$M_t(f_s) = \sum_{\alpha \in L^t} p_t(\alpha) \lceil -\log p(\alpha) \rceil$$

$$= -\sum_{\alpha \in L^t} p_t(\alpha) \log p_t(\alpha) - \Big(\log P(L^t) + O(1)\Big) \cdot \sum_{\alpha \in L^t} p_t(\alpha).$$

Поскольку

$$\log P(L^t) = O(\log t), \ \sum_{\alpha \in L^t} p_t(\alpha) = 1,$$

$$-\sum_{\alpha \in L^t} p_t(\alpha) \log p_t(\alpha) = H(\mathcal{L}^t) = O(t^2),$$

то справедливо равенство

$$C(f_s) = \lim_{t\to\infty} \frac{H(\mathcal{L}^t) + O(\log(t))}{\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha|} = C^*(\mathcal{L}).$$

Указанный алгоритм кодирования не является эффективным ввиду его большой вычислительной сложностью, которая обусловлена необходимостью упорядочивания слов языка по возрастанию вероятностей. Однако описанный в главе 3 эффективный алгоритм блочного кодирования f_{sh}, использованный в докритическом случае, будет асимптотически оптимальным и в данном случае при укрупнении грамматики $G(n)$ с $n\to\infty$, что доказывается аналогично неразложимому случае. Напомним вкратце идею доказательства. Избыточность кодирования f_{sh} может быть оценена как

$$\Delta(f_{sh}) = C(\mathcal{L}, f_{sh}) - C^*(\mathcal{L}) < \frac{\sum_{i,j} M(S_{ij}(t))\big(\lceil -logp_ij\rceil + \log p_ij\big)}{\sum_{\alpha\in L^t} p_t(\alpha)\cdot|\alpha|}$$

$$< \frac{\sum_{i\geq k_1} v''_{i-k_1}\sum_j p_{ij}(1+o(1))}{\sum_{i\geq k_1} v''_{i-k_1} L(R_i)}.$$

Учитывая нормировку $\sum_{i\geq k_1} v''_{i-k_1} = 1$ и условие $\sum_j p_{ij} = 1$, получаем, что

$$\Delta(f_{sh}) < \frac{1}{\sum_{i\geq k_1} v''_{i-k_1} L(R_i)}.$$

Избыточность кодирования f_{sh} при использовании укрупненной грамматики $G(n)$ можно оценить как

$$\Delta(f_{sh}, n) < \frac{1}{\sum_{i\geq k_1} v''_{i-k_1} L(R'_i)},$$

где

$$L(R'_i) = \sum_{j,q=1}^{k} M(S^i_{jq}(n)) l_{jq}.$$

Из этого неравенства и ранее доказанных оценок $M(S^i_{jq}(n)) = O(n^2)$ следует, что $\Delta(f_{sh}, n) \to 0$ при $t\to\infty$.

случае.

IV.2 Случай простого перронова корня

В этом параграфе будет исследована грамматика с двумя классами нетерминалов в случае, когда матрица первых моментов имеет вид (I.2.1), причем $r'\neq r''$, и $r = \max(r', r'') = 1$. Оказывается, что в этом случае, как и при $r<1$, асимптотика для $Q_i(t)$, $P_i(t)$, а также математических ожиданий числа применений правил $M(q_{ij}(t,\tau))$, $M(S_{ij}(t))$ на фиксированном ярусе τ дерева вывода из D^t_i и во всем дереве имеют такой же вид, как и в неразложимом случае. Доказательства проводятся аналогично неразложимому случаю [10], [16] с некоторыми небольшими отличиями, показанными далее. Случаи $r' < r''$ и $r' > r''$ рассматриваются отдельно, поскольку содержат некоторые отличия в доказательствах.

IV.2.1 Вероятности продолжения и вероятности деревьев вывода фиксированной высоты

Заметим, что в случае $r' < r''$ верны соотношения $v = (0, v'')$, $u > 0, v'' > 0$, а при $r' > r''$ - соотношения $u = (u', 0)$, $v > 0, u' > 0$. Для величины $R(t,s) = (R_1(t,s), \dots, R_k(t,s)) = 1 - F(t,s)$ справедливы следующие результаты.

Лемма IV.2.1.1 *Для грамматики с двумя классами нетерминалов, имеющей матрицу первых моментов вида (I.2.1) при* $r' < r'' = 1$ *верна оценка*

$$R_n(t,s) = \frac{u_n \sum\limits_{i=k_1+1}^{k} v_i \cdot (1 - s_i)}{1 + Bt \cdot \sum\limits_{i=k_1+1}^{k} v_i \cdot (1 - s_i)/2} \cdot (1 + o(1))$$

при $n = 1, \dots, k$*, где* B *задано формулой*

$$B = \sum_{i,j,l=1}^{k} v_i b_{jl}^{i} u_j u_l \tag{IV.2.1.1}$$

и $B > 0$*. Оценка является равномерной по* $0 \le s \le 1, s(k_1 + 1 : k) < 1$ *в области* $\max(1 - s)/\min_{j > k_1}(1 - s_j) < C$ *для заданной константы* $C > 0$

Этот результат доказывается точно так же, как и соответствующее утверждение в неразложимом случае [16]. Действительно, в равенстве

$$\bar{1} - F(t,s) = [A - E(F(t-1,s))][A - E(F(t-2,s))] \dots [A - E(1,s)](\bar{1} - s)$$

величины $A_t(s) = E_{jn}(F(t,s)) \le \sum_l b_{jn}^l \cdot (1 - F(l,0)) = A_t$ ограничены равномерно по s, причем $F_l(t,0) \to 1$ при $t \to \infty$. Поэтому равномерная сходимость

$$\frac{R_i(t,s)}{\sum\limits_{i=k_1+1}^{k} v_i R_i(t,s)} \to u_i$$

вытекает из лемм III.2.3 и III.2.4 и замечания в конце доказательства леммы III.2.3 о равномерности оценок в области $\max(1 - s)/\min_{j > k_1}(1 - s_j) < C$ и $A_t(s) < A_t$, где $A_t \to 0$ при $t \to \infty$.

Лемма IV.2.1.2 *Для грамматики с двумя классами нетерминалов, имеющей матрицу первых моментов вида (I.2.1), при* $1 = r' > r''$ *верны оценки*

$$R_n(t,s) = \frac{u_n \sum\limits_{i=1}^{k} v_i \cdot (1 - s_i)}{1 + Bt \cdot \sum\limits_{i=1}^{k} v_i \cdot (1 - s_i)/2} \cdot (1 + o(1))$$

при $n = 1, \dots, k_1$*, где* $B > 0$ *введено формулой (IV.2.1.1). Для* $n = k_1 + 1, \dots, k$ *верны оценки* $R_n(t,s) = O((r'')^t) \cdot \sum\limits_{i=k_1+1}^{k} v_i \cdot (1 - s_i)$*. Обе оценки выполнены равномерно по* $0 \le s \le 1,\ s \ne 1$.

Оценки для $R_n(t,s),\ n=1,\ldots,k_1$ доказываются таким же образом с помощью лемм III.2.3 и III.2.4, для $k_1<n\le k$ следуют из оценки $R(t,s)(k_1+1:k)\le (A^{(3)})^t(\bar 1-s)=O((r'')^t(\bar 1-s))$ для неразложимой грамматики, порожденной нетерминалами из K_2.

Из этих двух лемм сразу же получаются оценки для величин $Q_i(t)$ и $P_i(t)=Q_i(t-1)-Q_i(t)$. Соотношение $\frac{P_i(t)}{\sum_i v_iP_i(t)}\to u_i$ получается из леммы III.2.3 таким же образом, как и в неразложимом случае.

Сформулируем эти результаты в виде следующей теоремы.

Теорема IV.2.1.1 *Для грамматики с двумя классами нетерминалов, имеющей матрицу первых моментов вида (I.2.1) при $r'\ne r''$, $r=1$, для $i=1,\ldots,k$ верны асимптотические равенства*

$$Q_i(t)=\frac{2u_i+o(1)}{Bt},$$

$$P_i(t)=\frac{2u_i+o(1)}{Bt^2},$$

где B задано формулой (IV.2.1.1) и $B>0$.

Заметим, что $u_i=0$ при $r'>r''$ для $i>k_1$, поэтому указанные асимптотики в этом случае имеют вид $Q_i(t)=o(1/t),\ P_i(t)=o(1/t^2)$. Поскольку при $i>k_1$ величины $Q_i(t),P_i(t)$ совпадают с соответствующими величинами для неразложимой грамматики, порожденной нетерминалами из K_2, то при $r'>r''$ более точные асимптотики для $Q_i(t),P_i(t)$ имеют вид $O((r'')^t)$.

IV.2.2 Распределение нетерминалов на фиксированном ярусе дерева вывода

Для нахождения асимптотики величин $M(q_{ij}(t,\tau))$ нам необходимо исследовать распределение случайного вектора $x(t)=(x_1(t),\ldots,x_k(t))$ при условии $x(t)>0$, где $x_i(t), i=1,\ldots,k$ - число нетерминалов A_i на последнем ярусе в дереве вывода высоты t. Рассмотрим случайные величины $\xi^i_j(t)=\frac{2x^i_j(t)}{v_jBt}$, где $j=1,\ldots,k$ при $r'>r''$ и $j=k_1+1,\ldots,k$ при $r'<r''$, а $x^i_j(t)$ - количество нетерминалов A_i на ярусе t в деревьях вывода слов из L с корнем, помеченным нетерминалом A_j. Сформулируем теоремы, аналогичные теореме 6.3.3 из [16], для случаев $r'<r''$ и $r''<r'$.

Теорема IV.2.2.1 *Для грамматики с двумя классами нетерминалов, имеющей матрицу первых моментов вида (I.2.1) при $r'<r''=1$ условное распределение векторов $\xi''^i(t)=(\xi^i_{k_1+1}(t),\ldots,\xi^i_k(t))$, $i=1,\ldots,k_1$ при условии $\xi^i(t)=(\xi^i_1(t),\ldots,\xi^i_k(t))\ne 0$ сходится при $t\to\infty$ по распределению к случайному вектору $\xi''=(\xi''_{k_1+1},\ldots,\xi''_k)$, не зависящему от i, при этом*

$$P(\xi''_1<x)=1-e^{-x},\ x\ge 0,$$

и с вероятностью 1 справедливо равенство $\xi''_{k_1+1}=\xi''_{k_1+2}=\ldots=\xi''_k$.

Доказательство этого утверждения аналогично доказательству соответствующего утверждения для неразложимого случая. Рассмотрим преобразование Лапласа

$$\phi_t^i(\tau) = M\Big(\exp(-\sum_{j=k_1+1}^{k} \tau_j \xi_j''^i(t)|\xi^i(t) > 0)\Big) = 1 - \frac{R_i(t, s^\tau(t))}{Q_i(t)},$$

где $\tau = (\tau_{k_1+1}, \ldots, \tau_k)$, $s_j^\tau(t) = 0$ при $j \le k_1$, и $s_j^\tau(t) = \exp\big(-2\tau_j/(v_j Bt)\big)$ при $j > k_1$. Заметим, что $1 - s_j^\tau(t) = (2 + o(1))\tau_j/(v_j Bt)$ при $j > k_1$ ввиду оценки $e^{-x} = 1 - x + O(x^2)$, $x > 0$. Следовательно, существует константа C, такая что $(1 - s_i^\tau(t))/(1 - s_j^\tau(t)) < C$, $i, j = k_1 + 1, \ldots, k$ для любого t. Поэтому, применяя оценки для $R_i(t, s)$, полученные в лемме IV.2.1.1, получаем

$$R_n(t, s^\tau(t)) = \frac{2u_n\bar{\tau}/(Bt)}{1 + 2Bt\bar{\tau}/(2Bt)} \cdot (1 + \varepsilon(t)) = \frac{2u_n\bar{\tau}}{Bt(1 + \bar{\tau})} \cdot (1 + \varepsilon(t)),$$

где $\bar{\tau} = \tau_{k_1+1} + \ldots + \tau_k$, а $\varepsilon(t) \to 0$ при $t \to \infty$. Поэтому, используя оценку $Q_n(t) = (2 + o(1))u_n/Bt$, полученную в лемме IV.2.1.1, имеем

$$\lim_{t\to\infty} \phi_t^i(\tau) = 1/(1 + \bar{\tau}).$$

Поскольку полученная функция является преобразованием Лапласа для экспоненциального распределения, то, применяя результаты об однозначности соответствия между распределением и его преобразованием Лапласа (см. например [17]), получаем требуемое утверждение.

Для оценки количества нетерминалов из класса K_1 на ярусе дерева вывода нам потребуется следующая оценка для факториальных моментов $m_\alpha(t) = (m_\alpha^1(t), \ldots, m_\alpha^k(t))$ порядка α.

Лемма IV.2.2.1 *Для разложимой КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1), при $r' < r'' = 1$ верна оценка $m_\alpha^i(t) = O(_1(\bar{\alpha})(r')^t)$, если $i \le k_1$ и $\alpha_j \le k_1$ хотя бы для одного индекса j, и $m_\alpha^i(t) = O(_2(\bar{\alpha})t^{\bar{\alpha}-1})$, если $\alpha > k_1$. Здесь через $_{1,2}(\bar{\alpha})$ обозначена константа, не зависящая от t, но, возможно, разная для разных значений $\bar{\alpha}$*

Лемма доказывается индукцией по $n = \bar{\alpha}$. При $n = 1$ верны оценки $a_j^i(t) = O((r')^t)$ при $i, j \le k_1$ и $a_j^i(t) = O(1)$ при $j > k_1$, т.к $(A^{(1)})^t = O((r')^t)$, где $A^{(1)}$ - верхний левый блок матрицы первых моментов A, и $A^t = O((r'')^t) = O(1)$. Пусть утверждение доказано для некоторого n. Докажем его для $\bar{\alpha} = n + 1$.

Используем соотношение, доказанное в [16]:

$$m_\alpha^i(t) = \sum_{\tau=1}^{t} \sum_{s=1}^{k} a_s^i(t - \tau) z_\alpha^s(\tau - 1), \tag{IV.2.2.1}$$

где $z_\alpha^s(\tau - 1)$ является суммой конечного числа произведений моментов более низких порядков, причем

$$z_\alpha^s(t) = m_\alpha^s(t + 1) - \sum_n a_n^s m_\alpha^n(t). \tag{IV.2.2.2}$$

Используя оценки $a_j^i(t) = O(1)$ при $i \le k_1,\ j > k_1$, из этой формулы и оценок $m_{\alpha_1}^i(t) = O(t^{\bar{\alpha_1}})$ для $\alpha_1 > k_1,\ \bar{\alpha_1} < n+1$ сразу следует $m_\alpha^i(t) = O(t^n)$ для $\alpha > k_1,\ \bar{\alpha} = n+1$.

Из равенства (IV.2.2.2) вытекает, что $z_\alpha^s(t) = 0$, если $n > k_1$ и если $\alpha_j \le k_1$ для некоторого индекса j. Поэтому можно считать, что в этом случае $n \le k_1$. Используя оценки для $m_\alpha^i(t)$ при $\bar{\alpha} \le n$, получаем

$$m_\alpha^i(t) = O((r')^t) \cdot \sum_{\tau=1}^{t} \tau^{n-1}(r')^\tau = O((r')^t).$$

Пусть $x(t) = (x_1(t), \ldots, x_k(t))$ - случайная величина, где $x_i(t)$ - количество нетерминалов A_i на ярусе t в деревьях вывода слов языка. Из предыдущей леммы вытекает

Следствие IV.2.2.1 *При $i \le k_1$ справедливы оценки*

$$M(x_i(t)) = O(t(r')^t),\ P(x_i(t) > 0) = O(t \cdot (r')^t).$$

Первое утверждение сразу следует из предыдущей леммы, поскольку

$$M(x_i(t)) = a_i^1(t)/Q(t) = O(t(r')^t),\ M(x_i(t)^2) = \left(b_{ii}^1(t) + a_i^1(t)\right)/Q(t) = O(t \cdot (r')^t).$$

Второе утверждение следует из неравенства Чебышева $P(|x - M(x)| > \varepsilon) < D(x)/\varepsilon^2$, примененного к случайной величине $x_i(t)$ с $\varepsilon = 1 - M(x_i(t))$, поскольку для такого ε неравенство $|x_i(t) - M(x_i(t))| < \varepsilon$ эквивалентно условию $x_i(t) = 0$.

Это утверждение показывает, что в случае $r' < r''$ в векторе $x(t)$ пропорция $x_i(t)/x_j(t),\ i \le k_1, j > k_1$ стремится к нулю при $t \to \infty$.

Теорема IV.2.2.2 *Для разложимой грамматики, имеющей матрицу первых моментов вида (I.2.1) при $1 = r' > r''$, условное распределение векторов $\xi^i(t) = (\xi_1^i, \ldots, \xi_k^i),\ i = 1, \ldots, k_1$, при условии $\xi^i(t) \ne 0$ сходится при $t \to \infty$ по распределению к случайному вектору $\xi = (\xi_1, \ldots, \xi_k)$, не зависящему от i. При этом*

$$P(\xi_1 < x) = 1 - e^{-x},\ x \ge 0,$$

и с вероятностью 1 верно равенство $\xi_1 = \xi_2 = \ldots = \xi_k$.

Доказательство этого утверждения полностью повторяет доказательство аналогичного утверждения для неразложимого случая [16], поскольку в нем не используется существенным образом неразложимость, а только асимптотики для $R(t, s)$ и $Q(t)$, полученные в леммах IV.2.1.2 и IV.2.1.1 которые имеют такой же вид, как и в неразложимом случае.

IV.2.3 Математические ожидания числа применения правил в деревьях вывода

В этом параграфе мы получим оценки для $M(q_{ij}(t,\tau))$ и $M(S_{ij}(t))$ для случая $r' \neq r''$. Заметим, что асимптотики для $Q_i(t), P_i(t)$ имеют тот же вид, что и в неразложимом случае, и распределение нетерминалов на фиксированном ярусе дерева вывода имеет такой же вид. Поэтому теоремы 1,2,3 из [11] для неразложимого критического случая справедливы и в нашем случае, так как в них не используются предположения о неразложимости грамматики и строгой положительности собственных векторов, а нужны только асимптотики $Q_i(t), P_i(t)$ и утверждение о пропорциональном составе нетерминалов на ярусах дерева вывода (теоремы IV.2.2.1 и IV.2.2.2). Для случая $r' < r''$ влиянием первого нетерминала можно пренебречь ввиду следствия IV.2.2.1.

Поэтому при $r'' \neq r'$ асимптотики величин $M(q_{ij}(t,\tau))$ и $M(S_{ij}(t))$ такие же, как и в неразложимом случае.

Теорема IV.2.3.1 *Для разложимой стохастической КС-грамматики с матрицей первых моментов, имеющей вид (I.2.1) с перроновым корнем $r = 1$ и $r' \neq r''$, при $B > 0,\ t \to \infty$ выполняются асимптотические равенства:*

$$M(q_{ij}(t,\tau)) = (v_i + O(\varepsilon^2))p_{ij} \cdot B\tau(t-\tau)/t$$

при $\varepsilon t \leq \tau \leq (1-\varepsilon)t$ и

$$M(S_{ij}(t)) = p_{ij}v_iBt^2/6 + o(t^2),$$

где B введено формулой (IV.2.1.1).

В случае $r' < r''$ влиянием нетерминалов из K_1 можно пренебречь ввиду оценок моментов, полученных в лемме IV.2.2.1.

В случае $r' < r''$, когда $v(1:k_1) = 0$, могут быть получены более точные оценки для величин $M(q_{ij}(t,\tau))$ и $M(S_{ij}(t))$ при $i \leq k_1$.

Пользуясь леммой IV.2.2.1 и следствием IV.2.2.1, оценим величины $M(x_i(t,\tau)), M(q_{ij}(t,\tau))$ при $i \leq k_1$. Очевидно, число применений правила r_{ij} на ярусе τ меньше, чем $x_i(t)$. Поэтому

$$M(q_{ij}(t,\tau)) < M(x_i(t,\tau)) < \frac{p_{ij}}{P_1(t)} \cdot \sum_X P_X(\tau)x_i(\tau)R_X(t-\tau),$$

где $X = (x_1, \ldots, x_k)$. Доказанная для неразложимого критического случая в [11] оценка $R_X(t-\tau) = O(1/(t-\tau))$ верна и в нашем случае, поскольку при ее получении использовались только асимптотики для $Q_i(t), P_i(t)$. Поэтому справедливы оценки

$$M(x_i(t,\tau)) = O\left(t^2r'^{\tau}/(t-\tau)\right),\ M(q_{ij}(t,\tau)) = O\left(t^2r'^{\tau}/(t-\tau)\right).$$

Величина $M(S_{ij}(t))$ при $i \leq k_1,\ r' < r''$ может быть оценена по аналогии с доказательством теоремы IV.1.2.1 для случая кратного перронова корня.

Подставляя в соотношения (IV.1.2.7),(IV.1.2.9) оценки $Q_i(t) = o(1),\ Q_X(t) = 1 + o(1),\ P_i(t)/P_j(t) = u_i/u_j + o(1)$ и пользуясь формулой (IV.1.2.10) для вычисления

$R_X(t)$, получаем, что величины $\bar{M}_{ij}^q$ удовлетворяют следующим рекуррентным соотношениям при $i, j \leq k_1$:

$$\bar{M}_{ij}^q(t) = \delta_q^i p_{ij} \cdot (1 + o(1)) + \sum_l \sum_{m=1}^{k_1} p_{ql} s_{ql}^m \bar{M}_{ij}^m(t-1)(1 + o(1)) =$$

$$\delta_q^i p_{ij}(1 + o(1)) + \sum_{m=1}^{k_1} a_m^q \bar{M}_{ij}^m(t-1)(1 + o(1)), \qquad \text{(IV.2.3.1)}$$

аналогично для величин M_{ij}^q получаем:

$$M_{ij}^q(t) = \frac{1 + o(1)}{u_q} \cdot \left(\delta_q^i p_{ij} \cdot \sum_{m=1}^{k} s_{ij}^m u_m + \sum_l \sum_{m \leq k_1} p_{ql} s_{ql}^m \sum_{n=1}^{k} (s_{ql}^n - \delta_n^m) u_m \bar{M}'^m_{ij}(t-1) \right)$$

$$+ \sum_l \sum_{m=1}^{k_1} p_{ql} s_{ql}^m u_m M_{ij}^m(t-1)$$

$$= \frac{1 + o(1)}{u_q} \cdot \left(\delta_q^i p_{ij} S_{ij} + \sum_{m=1}^{k_1} G_m^q \bar{M}_{ij}^m(t-1) + \sum_{m \leq k_1} a_m^q u_m M_{ij}^m(t-1) \right), \qquad \text{(IV.2.3.2)}$$

где обозначено

$$S_{ij} = \sum_{m=1}^{k} s_{ij}^m u_m, \ G_m^q = \sum_{n=1}^{k} b_{mn}^q u_n.$$

Для решения этих рекуррентных соотношений нам потребуется следующая лемма :

Лемма IV.2.3.1 *Пусть $A_i \to A$ при $i \to \infty$ - последовательность матриц размером $k \times k$, причем матрица $A > 0$ и ее перронов корень $0 < r < 1$. Пусть $b_i \to b$ при $i \to \infty$ - последовательность векторов размером k, где $b \geq 0$, $b \neq 0$. Тогда для любого вектора x_0 последовательность $x_i, i = 0, 1, \ldots,$ определяемая рекуррентным соотношением $x_t = b_t + A_t x_{t-1}$, сходится к положительному вектору $x_* = (E - A)^{-1} b$.*

Идея доказательства такая же как в лемме IV.1.2.1. Докажем, что для любого $\varepsilon > 0$ существует номер n_0, такой, что $|x_* - x_n| < \varepsilon$ при $n > n_0$. Сначала докажем, что x_n ограничено некоторой константой C. Возьмем такой номер n_1, что $A_n < A/r_1$, $b_n < b/r_1$ для некоторого $r_1 > r, r_1 < 1$ при $n \geq n_1$. Используя обозначение $A^*(t, n) = A_t A_{t-1} \ldots A_{n+1}$ при $t > n$, и $A^*(t, n) = E$ при $t = n$, можем записать, что

$$x_{t+n_1} = A^*(t + n_1, n_1) x_{n_1} + \sum_{m=1}^{t} A^*(t + n_1, m + n_1) b_{m+n_1}. \qquad \text{(IV.2.3.3)}$$

Из этого соотношения сразу следует, что величины x_t ограничены, поскольку $A^*(t + n_1, m + n_1) = O((r/r_1)^{t-m})$ при $m < t$, то есть можно считать, что $x(t) < C$ для некоторой константы $C > 0$.

Матричный ряд $\sum\limits_{m=0}^{\infty} A^m$ сходится (поэлементно), поскольку его члены положительны и имеют порядок $O(r^m)$. Прямой проверкой показывается, что его сумма дает обратную матрицу к $E - A$ (которая в силу этого рассуждения положительна). Зададим произвольно малые величины $0 < \varepsilon_1, \varepsilon_2 < 1$. Возьмем такое m_0, что сумма остатка этого ряда $\sum\limits_{m=m_0}^{\infty} A^m < \varepsilon_1$, и $A^m/(r_1)^m < \varepsilon_1$ при $m \geq m_0$. Возьмем такое $n_2 > n_1$, что $|b_n - b| \leq \varepsilon_2 b$, и $|A_n - A| < \varepsilon_2 A$ при $n \geq n_2$. Тогда, пользуясь равенством (IV.2.3.3), получаем, что при $n \geq n_2$ справедливо неравенство

$$|x_{m_0+n} - x_*| = \Big| A^*(m_0+n,n)x_n + \sum_{m=0}^{m_0} \left(A^*(m_0+n,m+n)b_{m+n} - A^{m_0-m}b\right)$$
$$+ \left(\sum_{i=0}^{m_0} A^i b - (E-A)^{-1}b\right)\Big|.$$

Из неравенства $|A - A_n| < \varepsilon_2 A$ и положительности A_n при $n > n_2$ следуют оценки

$$|A^*(m_0+n,n)x_n| < A^{m_0}k \cdot (1+\varepsilon_2)^{m_0} \max x_n < kC \cdot (1+\varepsilon_2)^{m_0}\varepsilon_1,$$
$$\Big|\sum_{i=0}^{m_0} A^i b - (E-A)^{-1}b\Big| = \sum_{i>m_0} A^m b < k\max(b)\varepsilon_1,$$
$$|A^*(m_0+n,m+n)b_{m+n} - A^{m_0-m}b| \leq \left((1-\varepsilon_2)^{-m_0} - 1\right) \cdot A^{m_0-m}b,$$

Поэтому

$$|x_{m_0+n} - x_*| < kC \cdot (1+\varepsilon_2)^{m_0}\varepsilon_1 + \left((1-\varepsilon_2)^{-m_0} - 1\right) \cdot \max\Big(\sum_{i=0}^{m_0} A^i b\Big) + k\max(b)\varepsilon_1.$$

Поскольку элементы матриц $A^i = O(r^i)$ ограничены сверху некоторой константой C_1, то

$$|x_{m_0+n} - x_*| < \Big(C\varepsilon_1(1+\varepsilon_2)^{m_0(\varepsilon_1)} + \max(b)\varepsilon_1 + \left((1-\varepsilon_2)^{-m_0(\varepsilon_1)} - 1\right)m_0(\varepsilon_1)C_1\max(b)\Big)k.$$

Устремляя сначала ε_1, затем ε_2 к нулю, получаем соответствующие номера m_0, n_2, такие что $|x_* - x_{m_0+n}| < \varepsilon$ для любого сколь угодно малого $\varepsilon > 0$ при $n > n_2$. Лемма доказана.

Решая рекуррентные соотношения (IV.2.3.1) для $\bar{M}_{ij}^q(t)$ с использованием леммы IV.2.3.1, получаем, что при $i \leq k_1,\ t \to \infty$

$$\bar{M}_{ij}(t) \to (\bar{M}_{ij}^1(t), \ldots, \bar{M}_{ij}^{k_1}(t)) = p_{ij} \cdot (1+o(1)) \cdot (E - A^{(1)})^{-1}\delta^i = \bar{M}_{ij}^*,$$

и соответственно для величин $M_{ij}(t) = \left(u_1 M_{ij}^1(t), \ldots, u_{k_1} M_{ij}^{k_1}(t)t\right)$ из рекуррентных соотношений (IV.2.3.2) с помощью IV.2.3.1 находим асимптотику

$$M_{ij}(t) = p_{ij} \cdot (1+o(1)) \cdot \left(S_{ij} + \sum_m G_m^q \bar{M}_{ij}^{*m}\right) \cdot (E - A^{(1)})^{-1}\delta^i.$$

Таким образом, верна следующая:

Теорема IV.2.3.2 *Для КС-грамматики с матрицей первых моментов вида (I.2.1), при* $r' < r'' = 1,\ B > 0$*, где* B *дано формулой (IV.2.1.1), при* $i \le k_1$ *и* $t, \tau \to \infty$ *имеют место оценки:*

$$M(x_i(t,\tau)) < C_1 \cdot \Big(t^2 (r')^\tau / (t-\tau) \Big),$$

$$M(q_{ij}(t,\tau)) < C_2 \cdot \Big(t^2 (r')^\tau / (t-\tau) \Big).$$

для некоторых $C_1, C_2 > 0$*. Кроме того, для* $i \le k_1$ *при* $t \to \infty$

$$M(\bar{S}^q_{ij}(t)) \to p_{ij} \cdot (1 + o(1)) z^i_q,$$

$$M(S^q_{ij}(t)) \to p_{ij} \cdot \left(S_{ij} + \sum_m G^q_m p_{ij} z^i_m \right) z^i_q (1 + o(1))/u_q,$$

где обозначено

$$S_{ij} = \sum_{m=1}^{k} s^m_{ij} u_m,\ G^q_m = \sum_{n=1}^{k} b^q_{mn} u_n,\ z^i = (E - A^{(1)})^{-1} \delta^i.$$

Из этой теоремы видно, что в случае $0 < r' < r'' = 1$ в дереве вывода на каждом ярусе нетерминалов из класса K_2 значительно больше (O(t)), чем из K_1. Кроме того, общее число применений правил $r_{ij},\ i \le k_1$ в деревьях вывода высоты t стремится к константе.

IV.2.4 Алгоритм оптимального кодирования

Поскольку асимптотика для величин $M(S_{ij}(t))$ и $P(D^t)$ для случая $r' \neq r''$ такая же, как и в неразложимом случае, то выражения для энтропии $H(\mathcal{L}^t)$ множества слов $\mathcal{L}^t$ и нижней оценки стоимости оптимального кодирования $C^*(\mathcal{L})$ имеют такой же вид, как и в неразложимом критическом случае, причем схема кодирования, предложенная для случая кратного перронова корня в предыдущем разделе, является оптимальной и здесь, так как $H(\mathcal{L}^t) = O(t^2), P(t) = O(t^2)$. Эффективный алгоритм асимптотически оптимального при укрупнении правил грамматики кодирования, описанный в главе 3, будет асимптотически оптимальным и в этом случае.

Можно заметить, что в случае $r' < r''$ оценки для $H(\mathcal{L}^t)$ и $C^*(\mathcal{L})$ имеют тот же вид, что и для грамматики, порожденной только нетерминалами из K_2 (это следует из того, что в этом случае $v(1 : k_1) = 0$) - как и для случая кратного перронова корня.

Часть V

Заключение

А это заключение.

Список литературы

[1] Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Том 1. М.: Мир, 1978.

[2] Борисов А.Е. Кодирование слов стохастического КС- языка, порожденного разложимой грамматикой с двумя нетерминалами // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия Математика вып. 1(2), 2004. С. 18-28.

[3] Борисов А.Е. Закономерности в деревьях вывода для стохастической разложимой КС - грамматики.// Труды V международной научной конференции "Дискретные модели в теории управляющих систем". М.: Изд. отдел ВМиК МГУ, 2003. С. 15-17

[4] Борисов А.Е. О свойствах стохастического КС-языка, порожденного разложимой грамматикой.// Материалы научной конференции "Синтез и сложность управляющих систем". Н. Новгород, 2003. С. 15-18

[5] Борисов А.Е. О свойствах слов языка, порожденного разложимой стохастической КС-грамматикой с двумя нетерминалами. Критический случай.// Материалы VIII Международного семинара "Дискретная математика и ее приложения". М.: Изд. мех-мат. ф-та МГУ, 2004. С. 408-410.

[6] Борисов А.Е. О свойствах стохастического КС-языка, порожденного грамматикой с двумя классами нетерминальных символов.// Дискретный анализ и исследование операций. Серия 1, т.12, N3. Новосибирск: Издательство Института математики СО РАН, 2005. С.3-31.

[7] Борисов А.Е. О числе применений правил стохастической КС-грамматики.// Тезисы докладов XIV Международной конференции "Проблемы теоретической кибернетики". Изд. мех-мат. ф-та МГУ, 2005. С. 22.

[8] Гантмахер Ф.Р. Теория матриц. М.: Наука, 1967.

[9] Жильцова Л.П. Закономерности применения правил грамматики в выводах слов стохастического контекстно-свободного языка // Математические вопросы кибернетики. Вып.9. М.: Наука, 2000. С.101- 126.

[10] Жильцова Л.П. Закономерности в деревьях вывода слов стохастического контекстно-свободного языка и нижняя оценка стоимости кодирования. Критический случай.// Дискретный анализ и исследование операций. Серия 1, т.10, N3. Новосибирск: Издательство Института математики СО РАН, 2003. С.23-53.

[11] Жильцова Л.П. О нижней оценке стоимости кодирования и асимптотически оптимальном кодировании стохастического контекстно-свободного языка // Дискретный анализ и исследование операций. Серия 1, т.8, N3. Новосибирск: Издательство Института математики СО РАН, 2001. С.26- 45.

[12] Кричевский Р.Е. Сжатие и поиск информации. М.:Радио и связь, 1989.

[13] Марков А.А. Введение в теорию кодирования. М.: Наука, 1982.

[14] Марков А.А., Смирнова Т.Г. Алгоритмические основания обобщенно-префиксного кодирования.// Доклады АН СССР,т. 274 N4, С.790-793, 1984.

[15] Марков А.А. О неукоторых мерах сложности м эффективности в алфавитном кодировании. Матем. вопросы кибернетики вып. 6, с.348-352, М.: Наука, Физматгиз, 1996.

[16] Севастьянов В.А. Ветвящиеся процессы. М.: Наука, 1971.

[17] Феллер В. Введение в теорию вероятностей и ее приложения, том 1. - М.:Мир, 1984.

[18] Фихтенгольц Г.М. Основы математического анализа. Том 2. М.: Наука, 1968.

[19] Фу К. Структурные методы в распознавании образов. М.: Мир, 1977.

[20] Харрис Т. Теория ветвящихся случайных процессов. М.: Мир, 1966.

[21] Шеннон К. Математическая теория связи. М.: ИЛ, 1963.

[22] Яблонский С.В. Введение в дискретную математику. М.:Наука, 1986.

[23] Сагитов С., Ватутин В.А. Разложимый критический ветвящийся процесс с двумя типами частиц.// Вероятностные проблемы дискретной математики. Труды мат.инст. Стеклов. 177 (1986), стр. 3-20.

Докл.

[24] Сагитов С., Ватутин В.А. Разложимый критический ветвящийся процесс Беллмана-Харриса с двумя типами частиц. I.//Теор. Вероятности. 33(1988), N3, стр. 460-472.

[25] Сагитов С., Ватутин В.А. Разложимый критический ветвящийся процесс Беллмана-Харриса с двумя типами частиц.II.//Теор. Вероятности. 34(1989), N2, стр. 216-227.

[26] Borisov A.E. Optimal coding cost for stochastic CF- language induced by decomposable grammar.// VI International Conference on Mathematical Modeling/Book of abstracts, N.Novgorod, 2004. pp. 72.

[27] Ziv J.,Lempel A. Compression of individual sequences via variable-rate coding. // IEEE Trans.Inf.Theory IT-24,5 (Sept. 1978), p.530-536.

[28] Ziv J.,Lempel A. A universal algorithm for sequential data compression. IEEE Trans. Inf. Theory IT 23,3 (1977), p.337-343.

[29] Jorma Rissanen, Glen G.Langdon. Universal modeling and coding. // IEEE Transactions on Information Theory, V.21, N 1,pp 12-23,1981.

[30] Jorma Rissanen. Ggeneralized Kraft inequality and arithmetic coding. // IBM Journal Res.Develop, 1976. V.20, N3, p.198-203.

[31] Haffman D.A. A method for construction of minimum-redundancy codes// Proc. IRE 1952, V.40, N.10, p1098-1101.

[32] Zhiltsova L. On Entropy and Optimal Coding Cost for Stochastic Language.// Fundamenta Informaticae,V.36,pp.285-305,1998.