AofA'10 DMTCS proc. **AM**, 2010, 387–400

Counting Markov Types

Philippe Jacquet[1] Charles Knessl[2†] Wojciech Szpankowski[3‡]

[1] *INRIA, Rocquencourt, 78153 Le Chesnay Cedex, France.* `Philippe.Jacquet@inria.fr`
[2] *Dept. Math. Stat. & Compt. Sci., Univ. of Illinois at Chicago, Chicago, IL 60607-7045, U.S.A.* `knessl@uic.edu`
[3] *Department of Computer Science, Purdue University, West Lafayette, IN, USA.* `spa@cs.purdue.edu`

The method of types is one of the most popular techniques in information theory and combinatorics. Two sequences of equal length have the same type if they have identical empirical distributions. In this paper, we focus on Markov types, that is, sequences generated by a Markov source (of order one). We note that sequences having the same Markov type share the same so called *balanced frequency matrix* that counts the number of distinct pairs of symbols. We enumerate the number of Markov types for sequences of length n over an alphabet of size m. This turns out to be asymptotically equivalent to the number of the balanced frequency matrices as well as with the number of special *linear diophantine equations*, and also balanced directed multigraphs. For fixed m we prove that the number of Markov types is asymptotically equal to

$$d(m)\frac{n^{m^2-m}}{(m^2-m)!},$$

where $d(m)$ is a constant for which we give an integral representation. For $m \to \infty$ we conclude that asymptotically the number of types is equivalent to

$$\frac{\sqrt{2}m^{3m/2}e^{m^2}}{m^{2m^2}2^m\pi^{m/2}}n^{m^2-m}$$

provided that $m = o(n^{1/4})$ (however, our techniques work for $m = O(\sqrt{n})$). These findings are derived by analytical techniques ranging from multidimensional generating functions to the saddle point method.

Keywords: Markov types, integer matrices, linear diophantine equations, multidimensional generating functions, saddle point method

1 Introduction

The *method of types* is one of the most popular and useful techniques in information theory (e.g., source and channel coding) and combinatorics. Two sequences of equal length are of the same type if they have identical empirical distributions. The essence of the method of types was known for some time in probability and statistical physics. But only in the 1970's Csiszár and his group developed a general

† This author was supported by NSF Grant DMS-0800568 and NSA Grant H98230-08-1-0092
‡ The work of this author was supported by NSF STC Grant CCF-0939370, NSF Grants DMS-0800568, and CCF-0830140, and NSA Grant H98230-08-1-0092.

method and made it a basic tool of information theory of discrete memoryless systems [5]; see also [4, 6, 9, 11, 12, 14, 20, 21].

In this paper, we discuss Markov types. For concreteness, we first focus on Markov sources of order one. Let $\mathcal{A} = \{1, 2, \ldots, m\}$ be an m-ary alphabet, and consider a class of Markov distributions $\mathcal{P}_n(m)$ on $\mathcal{A}^n$, $n \geq 1$. We often simply write $\mathcal{P}_n := \mathcal{P}_n(m)$. Throughout, we study sequences $x^n = x_1 \ldots x_n$ of length n. For a distribution $P \in \mathcal{P}_n$ the *type class* of x^n is defined as

$$\mathcal{T}_n(x^n) = \{y^n : \ P(x^n) = P(y^n)\},$$

that is, all sequences having the same distribution as $P(x^n)$. Clearly, $\bigcup_{x^n} \mathcal{T}_n(x^n) = \mathcal{A}^n$. The *number* of type classes, or equivalently, the number of distinct distributions, is therefore equal to the cardinality of $\mathcal{P}_n(m)$ that we also often simplify to $|\mathcal{P}_n| := |\mathcal{P}_n(m)|$. We aim at deriving an asymptotic expression for $|\mathcal{P}_n|$ for fixed or large m when $n \to \infty$. For example, for binary memoryless sources, there are $|\mathcal{P}_n(2)| = n + 1$ type classes, and a class consisting of all sequences containing k 1's has cardinality $|\mathcal{T}_n(x^n)| = \binom{n}{k}$.

Markov sources X_t and corresponding Markov types are completely characterized by their transition probabilities that we denote by $\mathbf{P}$, that is $\mathbf{P} = \{p_{ij}\}_{i,j \in \mathcal{A}}$ is the transition matrix of $p_{ij} = P(X_{t+1} = j | X_t = i)$. The empirical distribution of a sequence x^n (with some fixed initial state) is

$$P(x^n) = \prod_{i,j \in \mathcal{A}} p_{ij}^{k_{ij}},$$

where k_{ij} is the number of pairs $(ij) \in \mathcal{A}^2$ in the sequence x^n. For example, $P(01011) = p_{01}^2 p_{10} p_{11}$. In passing we should point out that in the above p_{ij} can be viewed either as "a formal indeterminate" or, better, as the ratio of the number of pairs $(ij) \in \mathcal{A}^2$ to the length of the string (empirical distribution).

For a fixed transition matrix $\mathbf{P}$, the frequency matrix $\mathbf{k} = \{k_{ij}\}_{i,j \in \mathcal{A}}$ defines a *type class*. This matrix is an integer matrix satisfying two properties:

$$\sum_{i,j \in \mathcal{A}} k_{ij} = n - 1,$$

and additionally for any $i \in \mathcal{A}$ [11, 21]

$$\sum_{j=1}^{m} k_{ij} = \sum_{j=1}^{m} k_{ji} \pm \delta(x_1 = x_n), \quad \forall i \in \mathcal{A},$$

where $\delta(A) = 1$ when A is true and zero otherwise. The last property is called the *flow conservation property* and is a consequence of the fact that the number of pairs starting with symbols $i \in \mathcal{A}$ must equal to the number of pairs ending with symbol $i \in \mathcal{A}$ with the possible exception of the last pair. To avoid this exception, throughout we only consider *cyclic* strings in which the first element x_1 follows the last x_n. Thus, we consider integer matrices $\mathbf{k} = [k_{ij}]$ satisfying the following two properties

$$\sum_{i,j \in \mathcal{A}} k_{ij} = n, \tag{1}$$

$$\sum_{j=1}^{m} k_{ij} = \sum_{j=1}^{m} k_{ji}, \quad \forall\, i \in \mathcal{A}. \tag{2}$$

Such integer matrices $\mathbf{k}$ will be called *balanced frequency matrices*. We shall call (2) the "conservation law" equation. In this paper, we enumerate the number of Markov types $|\mathcal{P}_n|$ which is, for $n \to \infty$, asymptotically the same as the number of distinct balanced frequency matrices satisfying (1) and (2); we call the number of these solutions $\mathcal{F}_n(m)$.

Example. Let's first consider a binary Markov source. The balanced frequency matrix is of the following form

$$\mathbf{k} = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$$

where the nonnegative integers k_{ij} satisfy

$$\begin{aligned} k_{11} + k_{12} + k_{21} + k_{22} &= n, \\ k_{12} &= k_{21}. \end{aligned}$$

The above system of linear equations can be reduced to

$$k_{11} + 2k_{12} + k_{22} = n, \tag{3}$$

and the enumeration of Markov types over a binary alphabet reduces to finding the number of nonnegative solutions of (3). The answer is obviously

$$\begin{aligned} |\mathcal{P}_n| \sim \mathcal{F}_n(2) &= \sum_{k_{12}=0}^{\lfloor \frac{n}{2} \rfloor} (n - 2k_{12} + 1) \\ &= \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) (n - \left\lfloor \frac{n}{2} \right\rfloor + 1) = \frac{n^2}{4} + O(n). \end{aligned}$$

Let's now look at the $m = 3$ case. The balanced frequency matrix has nine elements $\{k_{ij}\}_{i,j \in \{1,2,3\}}$, and they satisfy

$$\begin{aligned} k_{11} + k_{12} + k_{13} + k_{21} + k_{22} + k_{23} + k_{31} + k_{32} + k_{33} &= n \\ k_{12} + k_{13} &= k_{21} + k_{31} \\ k_{12} + k_{32} &= k_{21} + k_{23} \\ k_{13} + k_{23} &= k_{31} + k_{32}. \end{aligned}$$

How many nonnegative integer solutions does the above system of linear equations have? We shall show that it is asymptotically $\frac{n^6}{12 \cdot 6!}$. ■

Our goal is to enumerate the number of Markov classes, that is, to find (asymptotically) the cardinality of $|\mathcal{P}_n|$. Our previous example demonstrated that this number is asymptotic, for $n \to \infty$, to the number of nonnegative integer solutions to the system of linear equations (1)-(2). Such an enumeration, for a general class of system of homogeneous diophantine equations, was investigated in Chap. 4.6 of Stanley's book [16] (cf. also [10]). Stanley developed a general theory to construct the associated generating function. However, ultimately only the denominator of this generating function is given in a semi-explicit form in [16], thus allowing the author to derive the growth rate of the number of integer solutions.

In this paper, we propose an approach based on previous work of Jacquet and Szpankowski [11] where analytic techniques such as multidimensional generating functions and the saddle point method were used. This allows us to derive precise asymptotic results. In particular, for fixed m we prove that the number of Markov types is asymptotically equal to

$$|\mathcal{P}_n| \sim d(m)\frac{n^{m^2-m}}{(m^2-m)!}, \quad n \to \infty,$$

where $d(m)$ is a constant for which we only found an integral representation.(i) For large $m \to \infty$ with $m^4 = o(n)$ we find that asymptotically the number of types is

$$|\mathcal{P}_n| \sim \frac{\sqrt{2}m^{3m/2}e^{m^2}}{m^{2m^2}2^m\pi^{m/2}}n^{m^2-m}.$$

However, our techniques also allow us to derive asymptotics for $m^2 = o(n)$. In passing we observe that the number of Markov types are often needed for minimax redundancy evaluation [1, 2, 11, 13, 15, 19].

Markov types were studied in a series of papers; see [11, 12, 20, 21]. However, existing literature mostly concentrates on finding the cardinality of a Markov type class, that is, $|\mathcal{T}(x^n)|$ with an exception of Martin et al. [12]. In particular, Whittle [21] already in 1955 computed $|\mathcal{T}(x^n)|$ for Markov chains of order one Regarding the number of types, it was known for a long while [4, 5, 6] that they grow polynomially, but only in [20] Weinberger et al. mentioned (without proof) that $|\mathcal{P}_n| = \Theta(n^{m^2-m})$. This was recently rigorously proved by Martin et al. in [12] for tree sources (that include Markov sources) for fixed m. However, the constant was never identified. We accomplish it here, as well as present asymptotic results for large m.

The paper is organized as follows. In Section 2 we formulate precisely our problem and present our main results for fixed and large m. The proofs are given in Section 3.

2 Main Results

In this section we present our main results and some of their consequences. We start with some notation. Throughout the paper, we let $\mathcal{F}$ be the set of all integer matrices $\mathbf{k}$ satisfying the conservation law equation (2), that is,

$$\sum_{j=1}^{m} k_{ij} = \sum_{j=1}^{m} k_{ji}, \quad \forall\, i \in \mathcal{A}.$$

For a given n, we let $\mathcal{F}_n$ be a subset of $\mathcal{F}$ consisting of matrices $\mathbf{k}$ such that the balance equations (1) and (2) hold.

We first make some general observations about generating functions over matrices, and summarize some results obtained in [11]. In general, let $g_{\mathbf{k}}$ be a sequence of scalars indexed by matrices $\mathbf{k}$ and define the generating function

$$G(\mathbf{z}) = \sum_{\mathbf{k}} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$$

(i) It is a simple exercise to extend these results to r order Markov chains. Indeed, one needs to replace m by m^r.

where the summation is over all integer matrices and $\mathbf{z} = \{z_{ij}\}_{i,j\in\mathcal{A}}$ is an $m \times m$ matrix that we often denote simply as $\mathbf{z} = [z_{ij}]$ (assuming the indices i and j run from 1 to m). Here $\mathbf{z}^{\mathbf{k}} = \prod_{i,j} z_{ij}^{k_{ij}}$ where k_{ij} is the entry in row i column j in the matrix $\mathbf{k}$. We denote by

$$G^*(\mathbf{z}) = \sum_{\mathbf{k}\in\mathcal{F}} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \sum_{n\geq 0}\sum_{\mathbf{k}\in\mathcal{F}_n} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$$

the $\mathcal{F}$-generating function of $g_{\mathbf{k}}$, that is, the generating function of $g_{\mathbf{k}}$ over matrices $\mathbf{k} \in \mathcal{F}$ satisfying the balance equations (1) and (2). The following useful lemma is proved in [11] but for completeness we repeat it here. Let $[z_{ij}\frac{x_i}{x_j}]$ be the matrix $\Delta^{-1}(x)\mathbf{z}\Delta(x)$ where $\Delta(x) = \mathrm{diag}(x_1, \ldots, x_m)$ is a diagonal matrix with elements $x_1, \ldots, x_m$, that is, the element z_{ij} in $\mathbf{z}$ is replaced by $z_{ij}x_i/x_j$.

Lemma 1 *Let $G(\mathbf{z}) = \sum_{\mathbf{k}} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$ be the generating function of a complex matrix $\mathbf{z}$. Then*

$$G^*(\mathbf{z}) := \sum_{n\geq 0}\sum_{\mathbf{k}\in\mathcal{F}_n} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \left(\frac{1}{2\mathbf{i}\pi}\right)^m \oint \frac{dx_1}{x_1}\cdots\oint\frac{dx_m}{x_m}G([z_{ij}\frac{x_j}{x_i}]) \tag{4}$$

with the convention that the ij-th coefficient of $[z_{ij}\frac{x_j}{x_i}]$ is $z_{ij}\frac{x_j}{x_i}$, and $\mathbf{i} = \sqrt{-1}$. In other words, $[z_{ij}\frac{x_j}{x_i}] = \Delta^{-1}(x)\mathbf{z}\Delta(x)$ where $\Delta(x) = \mathrm{diag}(x_1, \ldots, x_m)$. By the change of variables $x_i = \exp(\mathbf{i}\theta_i)$ we also have

$$G^*(\mathbf{z}) = \frac{1}{(2\pi)^m}\int_{-\pi}^{\pi} d\theta_1 \cdots \int_{-\pi}^{\pi} d\theta_m G([z_{ij}\exp((\theta_j - \theta_i)\mathbf{i})]$$

where $[z_{ij}\exp(\theta_j - \theta_i)] = \exp(-\Delta(\theta))\mathbf{z}\exp(\Delta(\theta))$.

Proof. Observe that

$$G(\Delta^{-1}(x)\mathbf{z}\Delta(x)) = G([z_{ij}\frac{x_j}{x_i}]) = \sum_{\mathbf{k}} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\prod_{i=1}^{m} x_i^{\sum_j k_{ji} - \sum_j k_{ij}}. \tag{5}$$

Therefore, $G^*(\mathbf{z})$ is the coefficient of $G([z_{ij}\frac{x_j}{x_i}])$ at $x_1^0 x_2^0 \cdots x_m^0$ since $\sum_j k_{ji} - \sum_j k_{ij} = 0$ for matrices $\mathbf{k} \in \mathcal{F}$. We write it in shortly as $G^*(\mathbf{z}) = \left[x_1^0 \cdots x_m^0\right] g([z_{ij}\frac{x_j}{x_i}])$. The result follows from the Cauchy coefficient formula (cf. [18]). ∎

We consider the number of solutions to (1) and (2) which is asymptotically equivalent to the number of Markov types $|\mathcal{P}_n(m)|$ over the alphabet $\mathcal{A}$, and whose generating function is

$$F_m^*(z) = \sum_{n\geq 0}\mathcal{F}_n(m)z^n.$$

Then applying the above lemma with $z_{ij} = zx_i/x_j$ we conclude that

$$F_m^*(z) = \frac{1}{(1-z)^m}[x_1^0 x_2^0 \cdots x_m^0]\prod_{i\neq j}\left[1 - z\frac{x_i}{x_j}\right]^{-1} \tag{6}$$

and thus, by the Cauchy formula,

$$\mathcal{F}_n(m) = [z^n]F_m^*(z) = \frac{1}{2\pi i}\oint \frac{F_m^*(z)}{z^{n+1}}dz.$$

In the next section we evaluate asymptotically this expression to yield the following main result of this paper

Theorem 1 (i) *For fixed m and $n \to \infty$ the number of Markov types is*

$$|\mathcal{P}_n(m)| = d(m)\frac{n^{m^2-m}}{(m^2-m)!} + O(n^{m^2-m-1}) \tag{7}$$

where $d(m)$ is a constant that also can be expressed by the following integral

$$d(m) = \frac{1}{(2\pi)^{m-1}} \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{(m-1)-fold} \prod_{j=1}^{m-1} \frac{1}{1+\phi_j^2}\cdot\prod_{k\neq\ell}\frac{1}{1+(\phi_k-\phi_\ell)^2}\, d\phi_1 d\phi_2\cdots d\phi_{m-1}. \tag{8}$$

(ii) *When $m \to \infty$ we find that*

$$|\mathcal{P}_n(m)| \sim \frac{\sqrt{2}m^{3m/2}e^{m^2}}{m^{2m^2}2^m\pi^{m/2}}\cdot n^{m^2-m} \tag{9}$$

provided that $m^4 = o(n)$.

Remark 1. It is easy to count the number of matrices $\mathbf{k}$ satisfying *only* equation (1), that is, $\sum_{ij} k_{ij} = n$. Indeed, in this case, it coincides with the number of integer solution of (1), which turns out to be the number of combinations with repetitions (the number of ways of selecting m^2 objects from n), that is,

$$\binom{n+m^2-1}{n} = \binom{n+m^2-1}{m^2-1} \sim n^{m^2-1}.$$

Thus the conservation law equation (2) decreases the above by the factor $\Theta(n^{m-1})$.

Remark 2. The evaluation of the integral (8) is quite cumbersome, but for small values of m we computed it to find that

$$|\mathcal{P}_n(2)| \sim \frac{1}{2}\frac{n^2}{2!} \tag{10}$$

$$|\mathcal{P}_n(3)| \sim \frac{1}{12}\frac{n^6}{6!} \tag{11}$$

$$|\mathcal{P}_n(4)| \sim \frac{1}{96}\frac{n^{12}}{12!} \tag{12}$$

$$|\mathcal{P}_n(5)| \sim \frac{37}{34560}\frac{n^{20}}{20!} \tag{13}$$

Tab. 1: Constants at n^{m^2-m} for fixed m and large m.

m	constant in (7)	constant in (9)
2	$1.920140832\,10^{-1}$	$2.500000000\,10^{-1}$
3	$9.315659368\,10^{-5}$	$1.157407407\,10^{-5}$
4	$1.767043356\,10^{-11}$	$2.174662186\,10^{-11}$
5	$3.577891782\,10^{-22}$	$4.400513659\,10^{-22}$

for large n. It appears that the coefficients of n^{m^2-m} are rational numbers, though we have no proof of this.

Remark 3. We now compare the coefficient at n^{m^2-m} for fixed m in (7) with its asymptotic counterpart in (9). They are shown in Table 1. Observe extremely small values of these constants even for relatively small m.

3 Analysis and Proofs

In this section we prove Theorem 1. Our starting formula is (6) that we repeat below

$$F_m^*(z) = \frac{1}{(1-z)^m}[x_1^0 x_2^0 \cdots x_m^0] \prod_{i \neq j} \left[1 - z\frac{x_i}{x_j}\right]^{-1}. \tag{14}$$

We first compute this explicitly for $m = 2, 3, 4, 5$.

When $m = 2$, we have

$$F_2^*(z) = \frac{1}{(1-z)^2}[x_1^0 x_2^0] \left[\frac{1}{1 - z\, x_1/x_2}\, \frac{1}{1 - z\, x_2/x_1}\right]. \tag{15}$$

Let us set $A = x_1/x_2$ so we need the coefficient of A^0 in $(1-Az)^{-1}(1-z/A)^{-1}$. Using a partial fractions expression in A, we have

$$\frac{1}{1-Az}\,\frac{1}{1-z/A} = \frac{1}{1-z^2}\left[\frac{1}{1-Az} + \frac{z}{A-z}\right].$$

For definiteness, we can assume that $|z| < |A| < |1/z|$ so that the coefficient of A^0 in $(1-Az)^{-1}$ is one and that in $z(A-z)^{-1}$ is zero. Hence, $F_2^*(z) = (1-z)^{-2}(1-z^2)^{-1} = (1+z)^{-1}(1-z)^{-3}$ and

$$|\mathcal{P}_n(2)| \sim \mathcal{F}_n(2) \quad = \quad \frac{1}{2\pi i}\oint \frac{1}{z^{n+1}}\frac{1}{1+z}\,\frac{1}{(1-z)^3}dz$$

$$= \frac{n^2}{4} + n + \frac{3}{4} + \frac{1}{8}[1 + (-1)^n] \sim \frac{1}{2}\,\frac{n^2}{2!}, \qquad n \to \infty. \tag{16}$$

For $m \geq 3$ we use recursive partial fractions expansions. When $m = 3$ we set $x_1/x_2 = A$, $x_1/x_3 = B$ so that we wish to compute

$$[A^0 B^0]\left(\frac{1}{1-zA}\,\frac{1}{1-z/A}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-Az/B}\,\frac{1}{1-Bz/A}\right). \tag{17}$$

First we do a partial fractions expansion in the A variable, for fixed B and z. Thus the factor inside the parentheses in (17) becomes

$$\begin{aligned} & \frac{1}{1-zA}\,\frac{1}{1-z^2}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-1/B}\,\frac{1}{1-Bz^2} \\ + & \frac{1}{1-z/A}\,\frac{1}{1-z^2}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-z^2/B}\,\frac{1}{1-B} \\ + & \frac{1}{1-Az/B}\,\frac{1}{1-B}\,\frac{1}{1-z^2/B}\,\frac{1}{1-B/z}\,\frac{1}{1-z/B}\,\frac{1}{1-z^2} \\ + & \frac{1}{1-Bz/A}\,\frac{1}{1-Bz^2}\,\frac{1}{1-1/B}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-z^2}. \end{aligned} \tag{18}$$

To coefficient of A^0 in the first term in (18) is

$$\frac{1}{1-z^2}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-1/B}\,\frac{1}{1-Bz^2}, \tag{19}$$

and that in the third term is

$$\frac{1}{1-B}\,\frac{1}{1-z^2/B}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-z^2}, \tag{20}$$

while the coefficients of A^0 are zero in the second and fourth terms. Combining (19) and (20) we must now compute

$$[B^0]\left(\frac{1+z^2}{1-z^2}\,\frac{1}{1-Bz}\,\frac{1}{1-z/B}\,\frac{1}{1-Bz^2}\,\frac{1}{1-z^2/B}\right). \tag{21}$$

Now expanding (21) by a partial fractions expansion in B leads to

$$\begin{aligned} & \frac{1+z^2}{1-z^2}[B^0]\;\left(\frac{1}{1-Bz}\,\frac{1}{1-z^2}\,\frac{1}{1-z}\,\frac{1}{1-z^2} + \frac{1}{1-z/B}\frac{1}{1-z^2}\,\frac{1}{1-z^3}\,\frac{1}{1-z}\right. \\ + & \left.\frac{1}{1-1/z}\,\frac{1}{1-z^3}\,\frac{1}{1-Bz^2}\,\frac{1}{1-z^4} + \frac{1}{1-z^3}\,\frac{1}{1-1/z}\,\frac{1}{1-z^4}\,\frac{1}{1-z^2/B}\right) \\ = & \frac{1+z^2}{1-z^2}\left[\frac{1}{1-z^2}\,\frac{1}{1-z}\,\frac{1}{1-z^3} + \frac{-z}{(1-z)}\,\frac{1}{1-z^3}\,\frac{1}{1-z^4}\right] = \frac{1-z+z^2}{(1-z)^4(1+z)^2(1+z+z^2)}. \end{aligned}$$

Hence,

$$F_3^*(z) = \frac{1-z+z^2}{(1-z)^7(1+z)^2(1+z+z^2)}.$$

Tab. 2: Poles and their orders for various m.

$m\backslash$ root	1	–1	$e^{\pm 2\pi i/3}$	$\pm i$	$e^{\pm 2\pi i/5}$	$e^{\pm 4\pi i/5}$
2	3	1	–	–	–	–
3	7	2	1	–	–	–
4	13	5	2	1	–	–
5	21	8	4	2	1	1

For $z \to 1$, $F_3^*(z) \sim \frac{1}{12}(1-z)^{-7}$ so that

$$|\mathcal{P}_n(3)| \sim \frac{1}{12}\frac{n^6}{6!}, \qquad n \to \infty. \tag{22}$$

Using similar recursive partial fractions expansions, with the help of the symbolic computation program MAPLE, we find that for $m = 4$ and $m = 5$

$$F_4^*(z) = \frac{z^8 - 2z^7 + 3z^6 + 2z^5 - 2z^4 + 2z^3 + 3z^2 - 2z + 1}{(1-z)^{13}(1+z)^5(1+z^2)(1+z+z^2)^2} \tag{23}$$

and

$$F_5^*(z) = \frac{Q(z)}{(1-z)^{21}(1+z)^8(1+z^2)^2(1+z+z^2)^4(1+z+z^2+z^3+z^4)}, \tag{24}$$

where

$$\begin{aligned} Q(z) &= z^{20} - 3z^{19} + 7z^{18} + 3z^{17} + 2z^{16} + 17z^{15} + 35z^{14} + 29z^{13} + 45z^{12} + 50z^{11} \\ &+ 72z^{10} + 50z^9 + 45z^8 + 29z^7 + 35z^6 + 17z^5 + 2z^4 + 3z^3 + 7z^2 - 3z + 1. \end{aligned}$$

These results show that it is unlikely that a simple formula can be found for $F_m^*(z)$ for general m.

By expanding (23) and (24) near $z = 1$ we conclude that as $n \to \infty$

$$|\mathcal{P}_n(4)| \sim \frac{1}{96}\frac{n^{12}}{12!}, \qquad |\mathcal{P}_n(5)| \sim \frac{37}{34560}\frac{n^{20}}{20!}. \tag{25}$$

It is easy to inductively show that at $z = 1$, $F_m^*(z)$ has a pole of order $m^2 - m + 1$ and the other singularities are poles at the roots of unity that are of order $< m^2 - m + 1$. These poles and their orders are given in Table 2.

Thus, for $n \to \infty$, we have

$$|\mathcal{P}_n(m)| \sim d(m)\frac{n^{m^2-m}}{(m^2-m)!}, \tag{26}$$

where

$$d(m) = \lim[(1-z)^{m^2-m+1}F_m^*(z)]$$

as $z \to 1$. However, there seems to be no simple formula for the sequence of constants $d(m)$. We proceed to characterize $d(m)$ as an $(m-1)$ fold integral.

First consider the simple case $m = 2$. Setting $A = e^{i\Phi}$ and using a Cauchy integral, we have

$$[A^0]\, \frac{1}{1-z/A}\, \frac{1}{1-Az} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{d\Phi}{1-2z\cos\Phi + z^2}.$$

Now set $z = 1-\delta$ and expand the integral for $z \to 1$. The major contribution will come from where $\delta \approx 0$ and scaling $\Phi = \delta\phi$ and using the Taylor expansion $1-2(1-\delta)\cos(\delta\phi)+(1-\delta)^2 = \delta^2[1+\phi^2]+O(\delta^3)$, we find that

$$F_2^*(z) \sim \frac{1}{\delta^2}\, \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\delta}{\delta^2[1+\phi^2]} d\phi = \frac{1}{2}\, \frac{1}{\delta^3}, \qquad \delta \to 0.$$

When $m = 3$, we use (3.4) and the Cauchy integral formula with $A = e^{i\Phi}$ and $B = e^{i\Psi}$ to get

$$\frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{1-2z\cos\Phi + z^2} \cdot \frac{1}{1-2z\cos\Psi + z^2} \cdot \frac{1}{1-2z\cos(\Phi-\Psi)+z^2}\, d\Phi d\Psi.$$

Again expanding the above for $z = 1-\delta \to 1$ and $\Phi = \delta\phi = O(\delta)$, $\Psi = \delta\psi = O(\delta)$, we obtain the leading order approximation

$$\frac{1}{\delta^4}\, \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{1+\phi^2}\, \frac{1}{1+\psi^2}\, \frac{1}{1+(\phi-\psi)^2}\, d\phi d\psi = \frac{1}{\delta^4} \cdot \frac{1}{12}.$$

Thus as $z \to 1$, $F_3^*(z) \sim \frac{1}{12}\delta^{-7} = \frac{1}{12}(1-z)^{-7}$ which follows also from the exact generating function.

For general m a completely analogous calculation shows that $F_m^*(z) \sim \delta^{m-m^2-1}d(m)$, as $\delta = 1-z \to 0$, where

$$d(m) = \frac{1}{(2\pi)^{m-1}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(m-1)-fold} \prod_{j=1}^{m-1} \frac{1}{1+\phi_j^2} \cdot \prod_{k \neq \ell} \frac{1}{1+(\phi_k-\phi_\ell)^2}\, d\phi_1 d\phi_2 \cdots d\phi_{m-1}. \tag{27}$$

We have verified that for $m = 4$ and $m = 5$, the integral agrees with our previous results. The second product in the above is over all distinct pairs (k,ℓ), so that this may also be written as

$$\prod_{\ell=1}^{m-2} \prod_{k=\ell+1}^{m-1} \frac{1}{1+(\phi_k-\phi_\ell)^2}. \tag{28}$$

This completes the proof of part (i) of Theorem 1.

We now use the saddle point method to prove part (ii) of Theorem 1. Since

$$\sum_{k_{ij}} \left(z\frac{z_i}{z_j}\right)^{k_{ij}} = (1 - z\frac{z_i}{z_j})^{-1}$$

and setting $z_i = e^{\mathbf{i}\theta_i}$ we find that

$$\begin{aligned} F_m^*(z) &= \frac{1}{(2\mathbf{i}\pi)^m} \oint \cdots \oint \prod_{ij} (1 - z\frac{z_i}{z_j})^{-1} \frac{dz_1}{z_1} \cdots \frac{dz_m}{z_m} & (29) \\ &= (2\pi)^{-m} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \prod_{ij} (1 - z\exp(i(\theta_i - \theta_j))^{-1} d\theta_1 \cdots d\theta_m \ . & (30) \end{aligned}$$

By noticing that the expression $\prod_{ij}(1 - z\exp(i(\theta_i - \theta_j))^{-1}$ does not change when the θ_i are all incremented of the same value, one can integrate over θ_1 to obtain

$$\begin{aligned} F_m^*(z) &= (2\pi)^{-m+1} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \prod_i (1 - z\exp(\mathbf{i}\theta_i))^{-1} (1 - z\exp(-\mathbf{i}\theta_i))^{-1} \\ &\times \frac{1}{1-z} \prod_{i>1, j>1} (1 - z\exp(\mathbf{i}(\theta_i - \theta_j))^{-1} d\theta_2 \cdots d\theta_m. & (31) \end{aligned}$$

Let now

$$\begin{aligned} L(z, \theta_2, \ldots, \theta_m) &= \log(1-z) + \sum_i \log(1 - z\exp(\mathbf{i}\theta_i))(1 - z\exp(-\mathbf{i}\theta_i)) \\ &+ \sum_{i>1, j>1} \log(1 - z\exp(\mathbf{i}(\theta_i - \theta_j)). \end{aligned}$$

An alternative form of the above is

$$\begin{aligned} L(z, \theta_2, \ldots, \theta_m) &= \log(1-z) + \sum_{i=2}^{m} \log(1 - 2z\cos\theta_i + z^2) \\ &+ \frac{1}{2} \sum_{i=2}^{m} \sum_{j=2}^{m} \log(1 - 2z\cos(\theta_i - \theta_j) + z^2). \end{aligned}$$

Notice that $L(z, 0, \ldots, 0) = m^2 \log(1-z)$. Hence

$$\mathcal{F}_n(m) = \frac{1}{\mathbf{i}(2\pi)^m} \oint \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp(-L(z, \theta_2, \ldots, \theta_m)) \frac{dz}{z^{n+1}} d\theta_2 \cdots d\theta_m. \quad (32)$$

In order to find the asymptotics of this integral we use the *multidimensional saddle point method* [18]. The quantity $L(z, \theta_2, \ldots, \theta_m) + n\log z$ attains its minimum value at $(\theta_2, \ldots, \theta_m) = (0, \ldots, 0)$ and $z = z_n = \frac{n}{m^2+n}$. The minimum value is therefore

$$m^2 \log(1 - z_n) + n\log z_n = m^2 \log(m^2/(m^2+n)) + n\log(n/(m^2+n))$$

or

$$m^2 \log(m^2) + n\log(n) - (m^2+n)\log(m^2+n).$$

Then

$$m^2 \log(1 - z_n) + n\log z_n = m^2 \log m^2 - m^2 \log n - m^2 + O(m^4/n)$$

provided that $m^4 = o(n)$.

After computations it turns out that at this point $(z, \theta_2, \ldots, \theta_m) = (z_n, 0, \ldots, 0)$:

$$\begin{aligned}
\frac{\partial^2}{\partial z^2} L(z, \theta_2, \ldots, \theta_m) &= -\frac{m^2}{(1-z_n)^2} \\
\forall i : \frac{\partial^2}{\partial z \partial \theta_i} L(z, \theta_2, \ldots, \theta_m) &= 0 \\
\forall i : \frac{\partial^2}{\partial \theta_i^2} L(z, \theta_2, \ldots, \theta_m) &= \frac{2(m-1)z_n}{(1-z_n)^2} \\
\forall i \neq j : \frac{\partial^2}{\partial \theta_i \partial \theta_j} L(z, \theta_2, \ldots, \theta_m) &= -\frac{2z_n}{(1-z_n)^2}, \quad m \geq 3.
\end{aligned}$$

In other words, the second derivative matrix $\mathbf{Q}_2$ of $L(z, \theta_2, \ldots, \theta_m) + n \log z$ at $(z, \theta_2, \ldots, \theta_m) = (z_n, 0, \ldots, 0)$ is

$$\mathbf{Q}_2 = \left(-\frac{m^2}{(1-z_n)^2} - \frac{n}{(z_n)^2} \right) \mathbf{u}_z \otimes \mathbf{u}_z + \frac{2mz_n}{(1-z_n)^2} \mathbf{I}_\theta - \frac{2z_n}{(1-z_n)^2} \mathbf{u}_\theta \otimes \mathbf{u}_\theta$$

where $\mathbf{u}_z = (1, 0, \ldots, 0)$,

$$\mathbf{u}_\theta = \frac{1}{\sqrt{m-1}} (0, 1, \ldots, 1),$$

and

$$\mathbf{I}_\theta = I - \mathbf{u}_z \otimes \mathbf{u}_z,$$

i.e., the identity restricted on θ components. In the above $\otimes$ is the tensor product (in our case, it is a product of two vectors resulting in a matrix). For example,

$$\mathbf{u}_\theta \otimes \mathbf{u}_\theta = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 1 & \ldots & 1 \end{pmatrix}.$$

An application of the saddle point method yields

$$\mathcal{F}_n(m) \sim \frac{1}{(2\pi)^{m/2} z_n \sqrt{\det(\mathbf{Q}_2)}} \exp(-m^2 \log(1-z_n) - n \log z_n)$$

where $\det(\cdot)$ denotes the determinant of a matrix. Since

$$|\det(\mathbf{Q}_2)| = \left(\frac{m^2}{(1-z_n)^2} + \frac{n}{(z_n)^2} \right) \left(\frac{z_n}{(1-z_n)^2} \right)^{m-1} 2^{m-1} m^{m-2} \sim n^{2m} m^{-3m} 2^{m-1},$$

we find that for $m^4 = o(n)$

$$|\mathcal{P}_n(m)| \sim \left(m^{-2m^2 + 3m/2} e^{m^2} 2^{-m} \pi^{-m/2} \sqrt{2} \right) n^{m^2 - m},$$

and this completes the proof. The condition $m^4 = o(n)$ is needed since we used the approximation for $m^2 \log(1 - z_n)$ below (32).

References

[1] K. Atteson, The Asymptotic Redundancy of Bayes Rules for Markov Chains, *IEEE Trans. on Information Theory*, 45, 2104-2109, 1999.

[2] A. Barron, J. Rissanen, and B. Yu, The Minimum Description Length Principle in Coding and Modeling, *IEEE Trans. Information Theory*, 44, 2743-2760, 1998.

[3] P. Billingsley, Statistical Methods in Markov Chains, *Ann. Math. Statistics*, 32, 12-40, 1961.

[4] T. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[5] I. Cszizar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[6] I. Cszizar, The Method of Types, *IEEE Trans. Information Theory*, 44, 2505-2523, 1998.

[7] L. D. Davisson, Universal Noiseless Coding, *IEEE Trans. Information Theory*, 19, 783-795, 1973.

[8] L. D. Davisson, Minimax Noiseless Universal coding for Markov Sources, *IEEE Trans. Information Theory*, 29, 211 - 215, 1983.

[9] M. Drmota and W. Szpankowski, Precise Minimax Redundancy and Regret, *IEEE Trans. Information Theory*, 50, 2686-2707, 2004.

[10] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2008.

[11] P. Jacquet and W. Szpankowski, Markov Types and Minimax Redundancy for Markov Sources, *IEEE Trans. Information Theory*, 50, 1393-1402, 2004.

[12] A. Martín, G. Seroussi, and M. J. Weinberger, Type classes of tree models, *Proc. ISIT 2007*, Nice, France, 2007.

[13] J. Rissanen, Fisher Information and Stochastic Complexity, *IEEE Trans. Information Theory*, 42, 40-47, 1996.

[14] G. Seroussi, On Universal Types, *IEEE Trans. Information Theory*, 52, 171-189, 2006.

[15] P. Shields, Universal Redundancy Rates Do Not Exist, *IEEE Trans. Information Theory*, 39, 520-524, 1993.

[16] R. Stanley, *Enumerative Combinatorics*, Vol. II, Cambridge University Press, Cambridge, 1999.

[17] W. Szpankowski, On Asymptotics of Certain Recurrences Arising in Universal Coding, *Problems of Information Transmission*, 34, 55-61, 1998.

[18] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.

[19] Q. Xie, A. Barron, Asymptotic Minimax Regret for Data Compression, Gambling, and Prediction, *IEEE Trans. Information Theory*, 46, 431-445, 2000.

[20] M. J. Weinberger, N. Merhav, and M. Feder, Optimal sequential probability assignment for individual sequences, *IEEE Trans. Inf. Theory*, 40, 384-396, 1994.

[21] P. Whittle, Some Distribution and Moment Formulæ for Markov Chain, *J. Roy. Stat. Soc.,* Ser. B., 17, 235-242, 1955.