Markov Types and Minimax Redundancy for Markov Sources*

March 27, 2004

Philippe Jacquet†
INRIA
Rocquencourt
78153 Le Chesnay Cedex
France
Philippe.Jacquet@inria.fr

Wojciech Szpankowski‡
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

Redundancy of universal codes for a class of sources determines by how much the actual code length exceeds the optimal code length. In the minimax scenario one designs the best code for the worst source within the class. Such minimax redundancy comes in two flavors: average minimax or worst case minimax. We study the *worst case* minimax redundancy of universal block codes for Markovian sources of any order. We prove that the maximal minimax redundancy for Markov sources of order r is asymptotically equal to $\frac{1}{2}m^r(m-1)\log_2 n + \log_2 A_m^r - (\ln\ln m^{1/(m-1)})/\ln m + o(1)$, where n is the length of a source sequence, m is the size of the alphabet and A_m^r is an explicit constant (e.g., we find that for a binary alphabet $m = 2$ and Markov of order $r = 1$ the constant $A_2^1 = 16\cdot G \approx 14.655449504$ where G is the Catalan number). Unlike previous attempts, we view the redundancy problem as an asymptotic evaluation of certain sums over a set of matrices representing Markov types. The enumeration of Markov types is accomplished by reducing it to counting Eulerian paths in a multigraph. In particular, we propose exact and asymptotic formulas for the number of strings of a given Markov type. All of these findings are obtained by analytic and combinatorial tools of analysis of algorithms.

Index terms: Minimax redundancy, Markov sources, Markov types, Eulerian paths, multidimensional generating functions, analytic information theory.

*A preliminary version of this paper was presented at *Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, Versailles, 2002.

†This work was partly supported by the Esprit Basic Research Action No. 7141 (Alcom II).

‡The work of this author was supported by the NSF Grants CCR-9804760 and CCR-0208709, and NIH grant R01 GM068959-01.

1 Introduction

In the *1997 Shannon Lecture* Jacob Ziv presented compelling arguments for "backing off" to a certain extent from first-order asymptotic analyses of information sources in order to predict the behavior of real systems with *finite* "description" length. One way of addressing this problem is to increase the accuracy of asymptotic analysis by replacing first-order analyses by full asymptotic expansions and more accurate analyses (for example, via large deviations or central limit laws). The *redundancy rate problem* in lossless source coding, which is the main topic of this paper, requires second-order asymptotics since one looks beyond the leading term of the code length. Thus, it is a perfect candidate for such studies. Recent years have seen a resurgence of interest in redundancy of lossless coding. Hereafter, we focus on redundancy of universal codes for Markov sources and present some precise asymptotic results.

To start, we introduce some definitions. A (block) code $C_n : \mathcal{A}^n \to \{0,1\}^*$ is defined as an injective mapping from the set $\mathcal{A}^n$ of all sequences of length n over the finite alphabet $\mathcal{A}$ of size $m = |\mathcal{A}|$ to the set $\{0,1\}^*$ of all binary sequences. We consider here only uniquely decipherable fixed-to-variable length codes. A source sequence of length n is denoted by $x_1^n \in \mathcal{A}^n$. We write X_1^n for a stochastic source producing a message of length n and $P(x_1^n)$ for the probability of generating x_1^n. For a given code C_n, we let $L(C_n, x_1^n)$ be the code length for x_1^n.

It is known that the entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n)\log P(x_1^n)$ is the absolute lower bound on the *expected* code length, where $\log := \log_2$ throughout the paper will denote the binary logarithm. Hence $-\log P(x_1^n)$ can be viewed as the "ideal" code length and therefore one may ask by how much the code length $L(C_n, x_1^n)$ exceeds the ideal code length, either for individual sequences or on average. The *pointwise redundancy* is

$$R_n(C_n, P; x_1^n) = L(C_n, x_1^n) + \log P(x_1^n),$$

while the *average redundancy* $\bar{R}_n(C_n, P)$ and the *maximal* redundancy $R_n^*(C_n, P)$ are defined, respectively, as

$$\begin{aligned} \bar{R}_n(C_n, P) &= \mathbf{E}_P[L(C_n, X_1^n)] - H_n(P), \\ R_n^*(C_n, P) &= \max_{x_1^n}[L(C_n, x_1^n) + \log P(x_1^n)], \end{aligned}$$

where the underlying probability measure P represents a particular source model and $\mathbf{E}$ denotes the expectation.

In practice, one can only hope to have some knowledge about a family of sources $\mathcal{S}$ that generates real data (e.g., memoryless sources $\mathcal{S} = \mathcal{M}_0$ or Markov sources of rth order $\mathcal{S} = \mathcal{M}_r$). Following Davisson [7] we define the average minimax redundancy $\bar{R}_n(\mathcal{S})$ and the worst case (maximal) minimax redundancy $R_n^*(\mathcal{S})$ for family $\mathcal{S}$, respectively, as follows

$$\bar{R}_n(\mathcal{S}) = \min_{C_n} \sup_{P \in \mathcal{S}} \left(\sum_{x_1^n} P(x_1^n)\left[L(C_n, x_1^n) + \log P(x_1^n)\right] \right), \tag{1}$$

$$R_n^*(\mathcal{S}) = \min_{C_n} \sup_{P \in \mathcal{S}} \max_{x_1^n} \left[L(C_n, x_1^n) + \log P(x_1^n)\right]. \tag{2}$$

That is, using either average minimax or worst case as our code evaluation criterion, we search for the best code for the worst source. We should also point out that there are other measures of optimality for coding such as *regret* functions defined as (cf. [2, 14, 23, 24])

$$\bar{r}_n(\mathcal{S}) = \min_{C_n} \sup_{P\in\mathcal{S}} \sum_{x_1^n} P(x_1^n)[L(C_n, x_1^n) + \log \sup_{P\in\mathcal{S}} P(x_1^n)]$$

but we shall not study these in the paper.

Our goal is to derive precise results for the worst case minimax redundancy $R_n^*(\mathcal{M}_r)$ for Markov sources $\mathcal{M}_r$ of order r. The *worst case* minimax redundancy is increasingly important since it measures the worst case excess of the best code maximized over the processes in a family of sources. In [14] Rissanen points out that the redundancy restricted to the first term cannot distinguish between codes that differ by a constant, however large; this constant can be large if the Fisher information of the data generating source is nearly singular. In this paper we pay special attention to the first two terms of the minimax redundancy for Markov sources.

To estimate the worst case minimax redundancy for any family of sources $\mathcal{S}$ we apply a recently derived formula [9] that improves the Shtarkov bound [16], namely

$$R_n^*(\mathcal{S}) = \log\left(\sum_{x_1^n} \sup_{P\in\mathcal{S}} P(x_1^n)\right) + R^{GS}(Q^*), \tag{3}$$

where $R^{GS}(Q^*)$ is the maximal redundancy of the generalized Shannon code (i.e., a code which assigns $\lceil \log 1/P(x_1^n)\rceil$ for some source sequences x_1^n and $\lfloor \log 1/P(x_1^n)\rfloor$ for remaining source sequences) designed for the maximal likelihood distribution

$$Q^*(x_1^n) = \frac{\sup_P P(x_1^n)}{\sum_{y_1^n} \sup_P P(y_1^n)}. \tag{4}$$

In $R_n^{GS}(Q^*)$ the distribution Q^* is assumed to be known. In passing we observe that the first part of (3) is a nondecreasing function of n depending *only* on the underlying class $\mathcal{S}$ of probability distributions, while the second term $R_n^{GS}(Q^*)$ contains a coding component and may be a fluctuating function of n.

For Markov sources $\mathcal{M}_r$ of order r, Drmota and Szpankowski [9] proved that the term $R_n^{GS}(Q^*)$ of $R_n^*(\mathcal{M}_r)$ is equal to

$$R_n^{GS}(Q^*) = -\frac{\ln\frac{1}{m-1}\ln m}{\ln m} + o(1). \tag{5}$$

Thus, hereafter we only deal with the first (leading) term of of $R_n^*(\mathcal{M}_r)$ that we denote as $\log D_n(\mathcal{M}_r)$, that is,

$$\log D_n(\mathcal{M}_r) = \log\left(\sum_{x_1^n} \sup_{P\in\mathcal{M}_r} P(x_1^n)\right).$$

We focus here on estimating asymptotically $D_n(\mathcal{M}_1)$ for Markov sources of order $r = 1$, and then generalize to any order r. In particular, we observe that

$$D_n(\mathcal{M}_1) = \sum_{\mathbf{k}} M_{\mathbf{k}} \left(\frac{k_{11}}{k_1}\right)^{k_{11}} \cdots \left(\frac{k_{m,m}}{k_m}\right)^{k_{m,m}},$$

where $k_i = \sum_{j=1}^m k_{ij}$ and $\mathbf{k} = \{k_{ij}\}_{i,j=1}^m$ is an integer matrix such that $\sum_{1\leq i,j\leq m} k_{ij} = n-1$. The quantity $M_{\mathbf{k}}$ denotes the number of strings x_1^n of type $\mathbf{k}$, that is, the number of strings x_1^n in which, for each $(i,j) \in \mathcal{A}^2$, symbol $i \in \mathcal{A}$ is followed by symbol $j \in \mathcal{A}$ a total of $k_{i,j}$ times. (Throughout the paper, we shall assume that $\mathcal{A} = \{1, 2, \ldots, m\}$ and write either $i \in \mathcal{A}$ or $a \in \mathcal{A}$.) Clearly, matrix $\mathbf{k}$ represents a Markovian type and $M_{\mathbf{k}}$ enumerates the number of strings belonging to the Markovian type $\mathbf{k}$.

In order to analyze $D_n(\mathcal{M}_1)$ we first need to estimate $M_{\mathbf{k}}$ asymptotically. This problem was previously studied by Whittle [25] (cf. [3, 4]), but we present here a novel approach based on generating functions and the enumeration of Eulerian paths in a multigraph. In particular, we prove that the number of strings $N_{\mathbf{k}}^{ba}$ starting with symbol a and ending with symbol b of type $\mathbf{k}$ is equal asymptotically to (cf. Theorem 1)

$$N_{\mathbf{k}}^{b,a} \sim \frac{k_{ba}}{k_b} \det_{bb}(\mathbf{I} - \mathbf{k}^*) \binom{k_1}{k_{11} \cdots k_{1m}} \cdots \binom{k_m}{k_{m1} \cdots k_{mm}}$$

where $\mathbf{k}^*$ is the matrix whose ij-th element is k_{ij}/k_i and $\det_{bb}(\mathbf{I} - \mathbf{k}^*)$ is the determinant of $(\mathbf{I} - \mathbf{k}^*)$ in which row b and column b are deleted.

The next step is to evaluate the sum in $D_n(\mathcal{M}_1)$. This sum turns out to fall into a special category that is worth studying on its own. Consider a matrix $\mathbf{k}$ as above with an addition property, called the *flow conservation property*,

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

(i.e., the sum of elements in the *ith* row is the same as the sum of elements in the ith column).[1] Let $\mathcal{F}_*$ be a set of all matrices $\mathbf{k}$ with the above property and $g_{\mathbf{k}}$ be a sequence indexed by $\mathbf{k}$. For our analysis it is crucial to find a relationship between the so called $\mathcal{F}$-generating function defined as

$$\mathcal{F}g(\mathbf{z}) = \sum_{\mathbf{k}\in\mathcal{F}_*} g_{\mathbf{k}} \mathbf{z}^{\mathbf{k}}$$

and the ordinary generating function

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

[1] We observe that a matrix $\mathbf{k}$ satisfying such an additional property is of Markovian type for *cyclic* strings in which the last symbol is followed by the first symbol. We shall discuss Markov types for cyclic strings in Section 2.2.

where the summation is over all integer matrices. In Lemma 1 we present a general approach to handle such sums.

Observe that $D_n(\mathcal{M}_1)$ is indeed a sum over $\mathcal{F}_*$. In our main results (cf. Theorem 2 and Theorem 3) we prove that

$$\log D_n(\mathcal{M}_r) = \frac{1}{2}m^r(m-1)\log n + \log A_m^r + O(1/n),$$

where A_m^r is an explicit constant. For example, we find that for a binary alphabet $m = 2$ and Markov of order $r = 1$ the constant $A_2^1 = 16 \cdot G \approx 14.655449504$ where G is the Catalan number.

Average and the worst case minimax redundancy have been studied since the seminal paper of Davisson [7]. Asymptotics of $R_n^*(\mathcal{M}_0)$ and $\bar{R}_n(\mathcal{M}_0)$ for memoryless sources have been known for some time (cf. [2, 12, 19, 23]). In fact, in [19] a full asymptotic expansion was derived. The leading term of the *average* minimax redundancy $\bar{R}_n(\mathcal{M}_r)$ for Markov sources $\mathcal{M}_r$ of order r was derived by Trofimov in [22] and subsequently improved by others. For example, Davisson proved in [8] that the second term of the average minimax redundancy is $O(1)$. Finally, recently Atteson [1] derived the two leading terms for the *average* minimax redundancy of $\mathcal{M}_r$ *ignoring* rounding the code length to an integer (i.e., ignoring in fact the coding part of the redundancy, as discussed above). There is, however, a lack of similar precise results for the worst case minimax redundancy for Markovian sources $\mathcal{M}_r$ of order r. Rissanen [14] obtained the first two terms of the worst case regret function, again ignoring rounding code lengths to integers (i.e., disregarding a term corresponding to $R_n^{GS}(Q^*)$ of (3)). Risannen's constant is expressed in terms of the Fisher information. In [11] lower and upper bounds for the worst case minimax redundancy were derived. In this paper we derive an asymptotic expansion of the worst case minimax redundancy for Markov sources of order r up to the constant term. However, the proposed methodology is, in principle, capable of producing a full asymptotic expansion for $R_n^*(\mathcal{M}_r)$. In [2, 9] the constant terms of the average and the maximal minimax redundancy are compared.

This paper is organized as follows. In the next section we present our main findings concerning Markov types and the worst case minimax redundancy. We derive these results in Section 3 using analytic tools of analysis of algorithms (cf. [21]). In passing we should point out that our goal is to obtain an asymptotic expansion of $R_n^*(\mathcal{S})$ for a large class of sources such as memoryless sources, Markov sources, mixing sources, and other non-parameterized class of sources. We aim at developing precise results of practical consequence using a combination of tools from average case analysis of algorithms, information theory, and combinatorics (cf. [9, 19, 21]).

2 Main Results

Following (3) and (5), we concentrate here on evaluating $D_n(\mathcal{M}_1)$ for Markov sources $\mathcal{M}_1$ of order one. We first compare $D_n(\mathcal{M}_1)$ to its corresponding formula $D_n(\mathcal{M}_0)$ for memoryless sources $\mathcal{M}_0$ over an m-ary alphabet. It is easy to see that $D_n(\mathcal{M}_0)$ is given by

$$D_n(\mathcal{M}_0) = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1,\ldots,k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}, \tag{6}$$

where k_i is the number of times symbol $i \in \mathcal{A}$ occurs in a string of length n. Indeed, we have

$$\sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} = \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}$$

and

$$\binom{n}{k_1,\ldots,k_m} = \frac{n!}{k_1!\cdots k_m!}$$

is equal to the number of strings x_1^n having k_i symbols $i \in \mathcal{A}$ (i.e, the number of strings in the type class $(k_1,\ldots,k_m)$). We present a brief analysis of (6) below in Section 2.1 as a preamble to our main analysis of Section 2.3 and Section 3.

Let us now turn our attention to the main topic of this paper, namely, the worst case minimax redundancy for Markov sources. We first focus on Markov sources of order $r = 1$. A similar argument to the one presented above yields

$$D_n(\mathcal{M}_1) = \sum_{\mathbf{k}} M_{\mathbf{k}} \left(\frac{k_{11}}{k_1}\right)^{k_{11}} \cdots \left(\frac{k_{m,m}}{k_m}\right)^{k_{m,m}}, \tag{7}$$

where $k_i = \sum_{j=1}^m k_{ij}$ and $\mathbf{k} = \{k_{ij}\}_{i,j=1}^m$ is an integer matrix[2] such that $\sum_{1\le i,j\le m} k_{ij} = n-1$. In the above, k_{ij} denotes the number of pairs $(i,j) \in \mathcal{A}^2$ in x_1^n, that is, the number of times symbol $j \in \mathcal{A}$ follows symbol $i \in \mathcal{A}$. The quantity $M_{\mathbf{k}}$ is the number of strings x_1^n generated over $\mathcal{A}$ having k_{ij} pairs (i,j) in x_1^n. It is known under the name *frequency count* (cf. [3]), but in fact it is the number of Markov strings of a given type. We call $\mathbf{k}$ the *pair occurrence* (PO) matrix for x_1^n or a Markovian type matrix.

2.1 Minimax Redundancy for Memoryless Sources

Let us first consider the class of memoryless sources $\mathcal{M}_0$ over an m-ary alphabet, that is, we shall study (6) for large n (and fixed m). In [19] we argued that such a sum can be analyzed through the so–called *tree generating function*. Let us define

$$B(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k = \frac{1}{1-T(z)}, \tag{8}$$

where $T(z)$ satisfies $T(z) = ze^{T(z)}$ and also $T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k$ (cf. [21]). Defining a new tree-like generating function, namely $D(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} D_k(\mathcal{M}_1)$, (6) and the convolution formula for generating functions (cf. [21]) immediately implies

$$D(z) = (B(z))^m.$$

[2] We sometimes abbreviate $\mathbf{k}$ by $[k_{ij}]$ to simplify some of our notation.

Let $[z^n]f(z)$ denote the coefficient of z^n in $f(z)$. Then, we finally arrive at

$$D_n(\mathcal{M}_0) = \frac{n!}{n^n}[z^n]\left(B(z)\right)^m .$$

To extract asymptotics from the above one must know the singular expansion of $B(z)$ around its singularity $z = e^{-1}$. But a minor modification of [5] gives

$$B(z) = \frac{1}{\sqrt{2(1-ez)}} + \frac{1}{3} - \frac{\sqrt{2}}{24}\sqrt{(1-ez)} + \frac{4}{135}(1-ez) - \frac{23\sqrt{2}}{1728}(1-ez)^{3/2} + O((1-ez)^2) .$$

Then an application of the Flajolet and Odlyzko *singularity analysis* [10] yields

$$\begin{aligned}\log D_n(\mathcal{M}_0) &= \frac{m-1}{2}\log\left(\frac{n}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + \frac{\Gamma(\frac{m}{2})m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\cdot\frac{\sqrt{2}}{\sqrt{n}} \\ &+ \left(\frac{3+m(m-2)(2m+1)}{36} - \frac{\Gamma^2(\frac{m}{2})m^2}{9\Gamma^2(\frac{m}{2}-\frac{1}{2})}\right)\cdot\frac{1}{n} + O\left(\frac{1}{n^{3/2}}\right)\end{aligned}$$

for large n.

2.2 Markov Types

In order to evaluate redundancy $D_n(\mathcal{M}_1)$ given by (7) for Markov sources $\mathcal{M}_1$ of order $r = 1$, we first need to estimate $M_{\mathbf{k}}$ for a given PO matrix $\mathbf{k}$. Since $\mathbf{k}$ can be viewed as a Markovian type, $M_{\mathbf{k}}$ is also the number of strings belonging to type $\mathbf{k}$. This problem was already addressed by Whittle [25]. Here we approach it from an analytic angle and derive, among others, asymptotics of $M_{\mathbf{k}}$.

First of all, we introduce the concept of *cyclic strings* in which the last symbol is followed by the first symbol. Observe that when we fix the first symbol of the string to $a \in \mathcal{A}$ and the last to $b \in \mathcal{A}$, then the PO matrix of such cyclic strings is simply $\mathbf{k}+[\delta_{ba}(i,j)]$, where we have used the Kronecker symbol notation in which $\delta_{ba}(i,j)$ is taken to be one if $(i,j) = (a,b)$ and zero otherwise. In the above, $\mathbf{k}$ is the PO matrix for regular strings. From now on we shall deal only with cyclic strings. Abusing slightly notation, we also write $\mathbf{k}$ for the PO matrix of cyclic strings. Observe that such matrices $\mathbf{k}$ satisfies the following two properties

$$\sum_{1\le i,j\le m} k_{ij} = n, \tag{9}$$

$$\sum_{j=1}^{m} k_{ij} = \sum_{j=1}^{m} k_{ji}, \qquad \forall\ i. \tag{10}$$

Property (10) is called the *conservation flow property*. From now on we assume that $\mathbf{k}$ satisfies (9)–(10).

Throughout the paper, we let $\mathcal{F}_*$ be the set of all integer matrices $\mathbf{k}$ satisfying property (10). For a given n, we let $\mathcal{F}_n$ be a subset of $\mathcal{F}_*$ consisting of matrices $\mathbf{k}$ such that $\sum_{ij} k_{ij} = n$, that is, (9) and (10) hold. For $\mathbf{k} \in \mathcal{F}_*$ we denote by $N_{\mathbf{k}}$ the number of cyclic strings of Markovian type $\mathbf{k}$. We also write $N_{\mathbf{k}}^a$ for the number of cyclic strings of type $\mathbf{k}$ starting

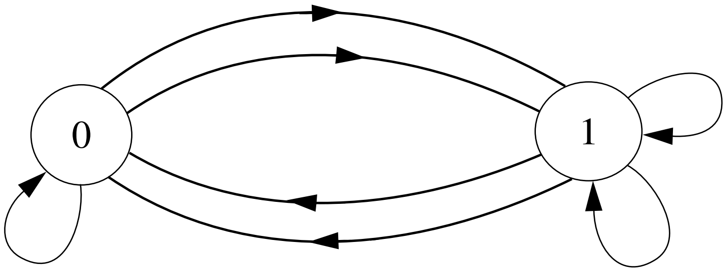

Figure 1: A directed multigraph for a binary alphabet $\mathcal{A} = \{0, 1\}$ with $k_{00} = 1$, $k_{01} = 2$, $k_{10} = 2$ and $k_{11} = 2$.

with a, and $N_{\mathbf{k}}^{ba}$ the number of cyclic strings of type $\mathbf{k}$ starting with a and ending with b; in other words, the number of cyclic strings of type $\mathbf{k}$ starting with ba.

We now reformulate the problem of enumerating cyclic strings with a given PO matrix $\mathbf{k}$ satisfying (9)-(10) as an enumeration problem on graphs. For a given matrix $\mathbf{k}$, let $\mathcal{G}_m$ be a directed multigraph defined on m vertices (labeled by the symbols from the alphabet $\mathcal{A} = \{1, 2, \dots, m\}$) with k_{ij} edges between the ith and jth vertex, where $i, j \in \mathcal{A}$. It is easy to see that the number of Eulerian paths starting with a vertex a is equal to $N_{\mathbf{k}}^{a}$. This is illustrated in Figure 1 for $\mathcal{A} = \{0, 1\}$, where the matrix $\mathbf{k}$ is

$$\mathbf{k} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}.$$

In order to present our first finding, we need to introduce some notation. Throughout, we shall use the following quantity

$$B_{\mathbf{k}} = \prod_{i \in \mathcal{A}} \frac{k_i!}{\prod_{j \in \mathcal{A}} k_{i,j}!} = \binom{k_1}{k_{11} \cdots k_{1m}} \cdots \binom{k_m}{k_{m1} \cdots k_{mm}} \tag{11}$$

where, we recall, $k_i = \sum_j k_{ij}$. Let also $\mathbf{z} = \{z_{ij}\}_{ij=1}^{m}$ be a complex $m \times m$ matrix and $\mathbf{k}$ an integer matrix. In the sequel, we write $\mathbf{z}^{\mathbf{k}} = \prod_{ij \in \mathcal{A}^2} z_{ij}^{k_{ij}}$. In particular, we have

$$B(\mathbf{z}) = \sum_{\mathbf{k}} B_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} = \prod_{a \in \mathcal{A}} (1 - \sum_{b \in \mathcal{A}} z_{a,b})^{-1}, \tag{12}$$

which is easy to check. We shall write $[\mathbf{z}^{\mathbf{k}}] f(\mathbf{z})$ for the coefficient of $f(\mathbf{z})$ at $\mathbf{z}^{\mathbf{k}}$ (e.g., $[\mathbf{z}^{\mathbf{k}}] B(\mathbf{z}) = B_{\mathbf{k}}$).

In Section 3.1 we prove the following result.

Theorem 1 *Let* $\mathbf{k} \in \mathcal{F}_n$ *for* $n \geq 1$.
(i) *For a given* $\mathbf{k}$, *the number* $N_{\mathbf{k}}^{a}$ *of cyclic strings of type* $\mathbf{k}$ *starting with symbol* a *is*

$$N_{\mathbf{k}}^{a} = [\mathbf{z}^{\mathbf{k}}] B(\mathbf{z}) \cdot \det_{aa}(\mathbf{I} - \mathbf{z}), \tag{13}$$

where $\mathbf{I}$ *is the identity matrix, and* $\det_{aa}(\mathbf{I} - \mathbf{z})$ *is the determinant of the matrix* $(\mathbf{I} - \mathbf{z})_{aa}$ *with the* a*-th column and the* a*-th row deleted.*

(ii) *The number* $N_{\mathbf{k}}^{ba}$ *of cyclic strings starting with the pair of symbols* ba *for which* $\mathbf{k}$ *is the PO matrix satisfies*

$$N_{\mathbf{k}}^{ba} = [\mathbf{z}^{\mathbf{k}}] z_{ba} B(\mathbf{z}) \cdot \det_{bb}(\mathbf{I} - \mathbf{z}). \tag{14}$$

Finally, as $n \to \infty$ *the frequency count* $N_{\mathbf{k}}^{ba}$ *attains the following asymptotics for* $k_{ba} = \Theta(n)$

$$N_{\mathbf{k}}^{b,a} = \frac{k_{ba}}{k_b} B_{\mathbf{k}} \cdot \det_{bb}(\mathbf{I} - \mathbf{k}^*) \left(1 + O\left(\frac{1}{n}\right)\right), \tag{15}$$

where $\mathbf{k}^*$ *is the matrix whose* ij*-th element is* k_{ij}/k_i*, that is,* $\mathbf{k}^* = [k_{ij}/k_i]$.

Remark. The enumeration of Eulerian paths in a multigraph is a classical problem (cf. [18]) and is related to the enumeration of spanning trees in a graph. Indeed, for a graph $\mathcal{G}_m$ built on m vertices with the adjacency matrix $\mathbf{k}$ we define the *Laplacian matrix* $\mathbf{L} = \{L_{ij}\}_{i,j \in \mathcal{A}}$ so that $L_{ij} = -k_{ij}$ for $i \neq j$ and $L_{ii} = \mathrm{outdeg}(i) - k_{ii}$, where $\mathrm{outdeg}(i)$ is the out-degree of vertex $i \in \mathcal{A}$. The *Matrix-Tree Theorem* [18] implies that the number $N_{\mathbf{k}}^{ba}$ of Eulerian paths in $\mathcal{G}_m$ with the first edge (ba) is given by

$$N_{\mathbf{k}}^{ba} = \det_{bb}(\mathbf{L}) \prod_{d \in \mathcal{A}} (\mathrm{outdeg}(d) - 1)!.$$

Equivalently, if $\mathbf{L}$ has m eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{m-1}, \lambda_m = 0$, then

$$N_{\mathbf{k}}^{ba} = \frac{1}{m} \lambda_1 \cdots \lambda_{m-1} \prod_{d \in \mathcal{A}} (\mathrm{outdeg}(d) - 1)!.$$

2.3 Minimax Redundancy for Markov Sources

In this section we formulate our main results concerning the worst case minimax redundancy for Markov sources. We start with a class $\mathcal{M}_1$ of Markov sources of order $r = 1$. We recall that the leading term $\log D_n(\mathcal{M}_1)$ of the minimax redundancy $R_n^*(\mathcal{M}_1)$ is given by (7). We re-write it for cyclic strings. First, we observe that $D_n(\mathcal{M}_1) = m D_n^a(\mathcal{M}_1)$ where the minimax redundancy $D_n^a(\mathcal{M}_1)$ is restricted to all strings starting with symbol a. Second, we recall that $N_{\mathbf{k}}^{ba}$ represents the number of cyclic strings starting with $a \in \mathcal{A}$ and ending with $b \in \mathcal{A}$, and with the PO matrix equal to $\mathbf{k}$. But $N_{\mathbf{k}}^{ba}$ is also the number of (regular) strings starting with $ba \in \mathcal{A}^2$ and with the PO matrix equal to $\mathbf{k} - [\delta_{ba}]$, where $[\delta_{ba}]$ is a matrix with all elements equal to 0 except the (ba)-th which is equal to 1.

We can now re-write our formula (7) for the redundancy of regular strings in terms of cyclic strings. We recall that $\mathbf{k}$ is the frequency matrix for cyclic strings and then $\mathbf{k} - [\delta_{ba}]$ is the frequency matrix for regular strings. Therefore, (7) becomes

$$D_n(\mathcal{M}_1) = m \sum_{b \in \mathcal{A}} \sum_{\mathbf{k} \in \mathcal{F}_n, k_{ba} > 0} N_{\mathbf{k}}^{ba} (\mathbf{k} - [\delta_{ba}])^{\mathbf{k} - [\delta_{ba}]} (k_b - 1)^{-k_b + 1} \prod_{i \neq b} (k_i)^{-k_i}, \tag{16}$$

where $\mathbf{k}^{\mathbf{k}} = \prod_{ij=1}^m k_{ij}^{k_{ij}}$. This formula is the starting point of our asymptotic analysis which is presented in full details in the next section.

In Section 3.2 we prove our second main results which is summarized next.

Theorem 2 *Let $\mathcal{M}_1$ be the class of Markov sources over a finite alphabet $\mathcal{A}$ of size m. The worst case minimax redundancy is*

$$R_n^*(\mathcal{M}_1) = \log D_n(\mathcal{M}_1) - \frac{\ln \frac{1}{m-1} \ln m}{\ln m} + o(1).$$

The leading term $D_n(\mathcal{M}_1)$ attains the following asymptotics as $n \to \infty$

$$D_n(\mathcal{M}_1) = \left(\frac{n}{2\pi}\right)^{m(m-1)/2} A_m \times \left(1 + O\left(\frac{1}{n}\right)\right) \tag{17}$$

with

$$A_m = m \int_{\mathcal{K}(1)} F_m(y_{ij}) \prod_{i\in\mathcal{A}} \frac{\sqrt{\sum_{j\in\mathcal{A}} y_{ij}}}{\prod_{j\in\mathcal{A}} \sqrt{y_{ij}}} \prod_{ij\in\mathcal{A}^2} dy_{ij}$$

where $\mathcal{K}(1) = \{y_{ij} : y_{ij} \geq 0,\ \sum_{ij} y_{ij} = 1, \forall i : \sum_j y_{ij} = \sum_j y_{ji}\}$, $F_m(\mathbf{y}) = \sum_{b\in\mathcal{A}} \det_{bb}(1-\mathbf{y}^*)$, *and $\mathbf{y}^*$ is the matrix whose ij-th coefficient is $y_{ij}/\sum_{j'} y_{ij'}$.*

We can evaluate the constant A_m for some small values of m. In particular, for a binary alphabet ($m = 2$) we have

$$A_2 = 2\int_{\mathcal{K}(1)} (\det_{11}(\mathbf{I}-\mathbf{y}^*) + \det_{22}(\mathbf{I}-\mathbf{y}^*)) \frac{\sqrt{y_1}}{\sqrt{y_{11}}\sqrt{y_{12}}} \frac{\sqrt{y_2}}{\sqrt{y_{21}}\sqrt{y_{22}}} dy_{11}dy_{12}dy_{21}dy_{22}. \tag{18}$$

Since $\det_{11}(\mathbf{I}-\mathbf{y}^*) = \frac{y_{21}}{y_2}$ and $\det_{22}(\mathbf{I}-\mathbf{y}^*)$ by symmetry, and since the condition $\mathbf{y} \in \mathcal{K}(1)$ means $y_1 + y_2 = 1$ and $y_{12} = y_{21}$ we arrive at, after the change of variable $x = \sin^2(\theta)$,

$$A_2 = 4\int_{y_{11}+2y_{12}+y_{22}=1} \frac{1}{\sqrt{y_{11}}\sqrt{y_1}\sqrt{y_{22}}\sqrt{y_2}} dy_{11}dy_{12}dy_{22}.$$

Therefore,

$$\begin{aligned} A_2 &= 4\int_0^1 \frac{dx}{\sqrt{(1-x)x}} \int_0^{\min\{x,1-x\}} \frac{dy}{\sqrt{(1-x-y)(x-y)}} \\ &= 8\int_0^{1/2} \frac{\log(1-2x) - \log(1-2\sqrt{(1-x)x})}{\sqrt{(1-x)x}} dx \\ &= 16\int_0^{\pi/4} \log\left(\frac{\cos(2\theta)}{1-\sin(2\theta)}\right) d\theta \\ &= 16 \cdot G, \end{aligned}$$

where G is the Catalan constant defined as $G = \sum_i \frac{(-1)^i}{(2i+1)^2} \approx 0.915965594$.

Next, we extend Theorem 2 to Markov sources of order r. A sketch of the proof is presented in Section 3.4.

Theorem 3 *Let $\mathcal{M}_r$ be the class of Markov sources of order r over a finite alphabet $\mathcal{A}$ of size m. The worst case minimax redundancy is*

$$R_n^*(\mathcal{M}_r) = \log D_n(\mathcal{M}_r) - \frac{\ln \frac{1}{m-1} \ln m}{\ln m} + o(1).$$

The leading term $D_n(\mathcal{M}_r)$ attains the following asymptotics as $n \to \infty$

$$D_n(\mathcal{M}_r) = \left(\frac{n}{2\pi}\right)^{m^r(m-1)/2} A_m^r \times \left(1 + O\left(\frac{1}{n}\right)\right) \tag{19}$$

with

$$A_m^r = m^r \int_{\mathcal{K}_r(1)} F_m^r(\mathbf{y}) \prod_{w \in \mathcal{A}^r} \frac{\sqrt{y_w}}{\prod_j \sqrt{y_{w,j}}},$$

where $\mathcal{K}_r(1)$ is the convex set of $m^r \times m$ matrices $\mathbf{y}$ with non-negative coefficients such that $\sum_{w,j} y_{w,j} = 1$, $w \in \mathcal{A}^r$. The function $F_m^r(\mathbf{y}_r) = \sum_w \det_{ww}(\mathbf{I} - \mathbf{y}_r^*)$, where $\mathbf{y}_r^*$ is the $m^r \times m^r$ matrix whose (w, w') coefficient is equal to $y_{w,a}/\sum_{i \in \mathcal{A}} y_{wi}$ if there exist a in $\mathcal{A}$ such that w' is a suffix of wa, otherwise the (w, w')th coefficient is equal to 0.

3 Analysis and Proofs

In this section we prove our main findings Theorem 1–3. The main methodological novelty of our approach lies in analytical treatment of certain sums over matrices satisfying the conservation flow property.

3.1 A Useful Lemma

In our setting the derivation of the minimax redundancy for Markov sources is reduced to the evaluation of a sum over the set $\mathcal{F}_*$ of matrices $\mathbf{k}$ satisfying (9)–(10). We need a method of handling such sums which is discussed next.

Let $g_{\mathbf{k}}$ be a sequence of scalars indexed by matrices $\mathbf{k}$ and

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

be its regular generating function. We denote by

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

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 (9)–(10).

The following lemma is useful. To write it in a compact form we introduce a short notation for matrices, namely, we shall write $[z_{ij}\frac{x_i}{x_j}]$ for 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 $x_1, \ldots, x_m$.

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*

$$\mathcal{F}g(\mathbf{z}) := \sum_{n \geq 0} \sum_{\mathbf{k} \in \mathcal{F}_n} g_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} = \left(\frac{1}{2\mathbf{i}\pi}\right)^m \oint \frac{dx_1}{x_1} \cdots \oint \frac{dx_m}{x_m} g([z_{ij}\frac{x_j}{x_i}]) \tag{20}$$

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 change of variable* $x_i = \exp(\mathbf{i}\theta_i)$ *we also have*

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

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

Proof. Observe that

$$g(\Delta^{-1}(x)\mathbf{z}\Delta(x)) = g([z_{ij}\frac{x_j}{x_i}]) = \sum_{\mathbf{k}} g_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\prod_{i=1}^{m} x_i^{\sum_i k_{ij}-\sum_j k_{ij}} \qquad (21)$$

Therefore, $\mathcal{F}g(\mathbf{z})$ is the coefficient of $g([z_{ij}\frac{x_j}{x_i}])$ at $x_1^0x_2^0\cdots x_m^0$ since $\sum_j k_{ij} - \sum_i k_{ij} = 0$ for matrices $\mathbf{k}\in\mathcal{F}_*$. We write it in shortly as $\mathcal{F}g(\mathbf{z}) = [x_1^0\cdots x_m^0]\; g([z_{ij}\frac{x_j}{x_i}])$. The result follows from the Cauchy coefficient formula (cf. [21]). ∎

Remark. Observe that (20) still holds when $g([z_{ij}x_j/x_i])$ is replaced by $g([z_{ij}x_i/x_j])$. We use this throughout the paper without any warning.

In particular, consider the sequence $B_{\mathbf{k}}$ defined in (11) whose generating function derived in (12) is recalled below

$$B(\mathbf{z}) = \sum_{\mathbf{k}} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \prod_{a\in\mathcal{A}}(1-\sum_{b\in\mathcal{A}} z_{a,b})^{-1}.$$

The generating function $\mathcal{F}B(\mathbf{z}) = \sum_{\mathbf{k}\in\mathcal{F}_*} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$ presented next is basically due to Whittle [25].

Corollary 1 *We have* $\mathcal{F}B(\mathbf{z}) = (\det(\mathbf{I}-\mathbf{z}))^{-1}$, *where* $\mathbf{I}$ *is the identity* $m\times m$ *matrix.*

Proof: For completeness we give our proof of Whittle's result. Setting $g_{\mathbf{k}} = B_{\mathbf{k}}$ in Lemma 1 and denoting $\mathbf{a} = \mathbf{I}-\mathbf{z}$ we find

$$\mathcal{F}B(\mathbf{z}) = \left(\frac{1}{2\mathbf{i}\pi}\right)^m \oint dx_1\cdots\oint dx_m \prod_i\left(\sum_j a_{ij}x_j\right)^{-1} = (\det(\mathbf{a}))^{-1} \qquad (22)$$

provided that $\mathbf{a}$ is not singular matrix. Indeed, one makes the linear change of variables $y_i = \sum_j a_{ij}x_j$ to obtain

$$\begin{aligned}\left(\frac{1}{2\mathbf{i}\pi}\right)^m \oint dx_1\cdots\oint dx_m \prod_i\left(\sum_j a_{ij}x_j\right)^{-1} &= (\det(\mathbf{a}))^{-1}\left(\frac{1}{2\mathbf{i}\pi}\right)^m\oint\frac{dy_1}{y_1}\cdots\oint\frac{dy_m}{y_m}\\ &= (\det(\mathbf{a}))^{-1}\end{aligned}$$

which completes the proof. ∎

Remark. Throughout this paper we also write $B_{\mathcal{A}}(\mathbf{z}) = \mathcal{F}_{\mathcal{A}}B(\mathbf{z})$ to simplify some notation where the subscript $\mathcal{A}$ indicates that the underlying alphabet is $\mathcal{A}$. In particular, from the above corollary one concludes that $B_{\mathcal{A}-\{a\}}(\mathbf{z}) = (\det_{aa}(\mathbf{I}-\mathbf{z}))^{-1}$, where $\det_{ij}(\mathbf{a})$ is the determinant of a matrix with the ith row and the jth column of $\mathbf{a}$ deleted.

For the proof of Theorem 2 we also need a continuous version of Lemma 1, which we establish next. Let $\mathcal{K}(x)$ be a hyper-polygon (simplex) of matrices $\mathbf{y}$ with non-negative *real* coefficients that satisfy the conservation flow property and such that $\sum_{ij} y_{ij} = x$. Observe that $\mathcal{F}_n$ is the set of non-negative integer matrices $\mathbf{k}$ that belongs to $\mathcal{K}(n)$. Let $a(x)$ be the area (hyper-volume) of $\mathcal{K}(x)$.

Lemma 2 *Let $g(\mathbf{x})$ be a function of real matrices $\mathbf{x}$. Let $G(\mathbf{t})$ be the Laplace transform of $g(\cdot)$, that is,*

$$G(t) = \int g(\mathbf{x}) \exp\left(-\sum_{ij} t_{ij} x_{ij}\right) d\mathbf{x},$$

and let

$$\tilde{G}(\mathbf{t}) = \int_0^\infty dy \int_{\mathcal{K}(y)} g(\mathbf{x}) \exp\left(-\sum_{ij} t_{ij} x_{ij}\right) d\mathbf{x}.$$

We have

$$\tilde{G}(\mathbf{t}) = \frac{1}{(2\mathbf{i}\pi)^m} \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta_1 \cdots \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta_m G([t_{ij} + \theta_i - \theta_j]) \tag{23}$$

where $[t_{ij} + \theta_i - \theta_j]$ is a matrix whose the ij-th coefficient is $t_{ij} + \theta_i - \theta_j$.

3.2 Markov Types and Eulerian Paths

We now prove Theorem 1. We recall that we evaluate the number $N_{\mathbf{k}}$ of cyclic strings of type $\mathbf{k}$. We start with a recollection of some definitions. Hereafter, we shall only deal with cyclic strings. In a cyclic string the first symbol follows the last one. If x is a cyclic string, then $k_{ij}(x)$ is the number of positions in x where symbol $j \in \mathcal{A}$ follows symbol $i \in \mathcal{A}$. The matrix $\mathbf{k} = \{k_{ij(x)}\}_{i,j=1}^m$ is the *pair occurrence* (PO) matrix for x. The PO matrix obviously satisfies the *conservation flow property* defined in (10). It is clear that in cyclic strings we have one pair occurrence more than in linear strings that results in $\sum_{ij} k_{ij} = n$, where n is the length of the string.

The key quantities of interest, called the *frequency counts*, are $N_{\mathbf{k}}$, $N_{\mathbf{k}}^a$, and $N_{\mathbf{k}}^{ba}$ for a given type (matrix) $\mathbf{k} \in \mathcal{F}_n$. We recall their definitions below:

- The frequency count $N_{\mathbf{k}}$ is the number of cyclic strings of type $\mathbf{k}$;
- $N_{\mathbf{k}}^a$ is the number of cyclic strings of type $\mathbf{k}$ starting with a symbol $a \in \mathcal{A}$;
- $N_{\mathbf{k}}^{b,a}$ is the number of cyclic strings of type $\mathbf{k}$ starting with a pair of symbols $ba \in \mathcal{A}^2$.

Notice that the frequency count $N_{\mathbf{k}}^{ba}$ is important for linear (regular) strings since it gives the number of strings starting with symbol a and ending with symbol b as a function of the PO matrix $\mathbf{k}$. Indeed, we know that one occurrence of the pair (ba) has to be removed from a cyclic string to make it a linear string.

3.2.1 Proof of Theorem 1(i)

Now we are in position to prove Theorem 1(i). We establish it in three separate steps. We first recall that $B_{\mathcal{A}}(\mathbf{z}) = \mathcal{F}B(\mathbf{z})$ with $B(\mathbf{z})$ defined in (12). By Corollary 1 we also know that $\mathcal{F}_{\mathcal{A}-\{a\}}B(\mathbf{z}) = B_{\mathcal{A}-\{a\}}(\mathbf{z}) = \det_{aa}^{-1}(\mathbf{I}-\mathbf{z})$ where $\det_{aa}(\mathbf{I}-\mathbf{z})$ is the determinant of the matrix $(\mathbf{I}-\mathbf{z})$ with row a and column a deleted.

We recall that we must prove that *for $n \geq 1$ and $\mathbf{k} \in \mathcal{F}_n$ the frequency count $N^a_{\mathbf{k}}$ is the coefficient at $\mathbf{z}^{\mathbf{k}}$ of $\frac{B(\mathbf{z})}{B_{\mathcal{A}-\{a\}}(\mathbf{z})}$, that is,*

$$N^a_{\mathbf{k}} = [\mathbf{z}^{\mathbf{k}}]\frac{B(\mathbf{z})}{B_{\mathcal{A}-\{a\}}(\mathbf{z})} = [\mathbf{z}^{\mathbf{k}}]B(\mathbf{z}) \cdot \det_{aa}(\mathbf{I}-\mathbf{z}) \tag{24}$$

where $B_{\mathcal{A}-\{a\}}(\mathbf{z})$ is the generating function of $B_{\mathbf{k}}$ over $\mathcal{A}-\{a\}$ satisfying the conservation flow property.

The proof proceeds via the enumeration of Euler cycles (paths) in a directed multigraph $\mathcal{G}_m$ over m vertices defined in the previous section. We recall that in such a graph vertices are labeled by symbols from the alphabet $\mathcal{A}$ with the edge multiplicity given by the matrix $\mathbf{k}$: there are k_{ij} edges from vertex $i \in \mathcal{A}$ to $j \in \mathcal{A}$. The number of Eulerian paths starting from vertex $a \in \mathcal{A}$ in a such multigraph is equal to $N^a_{\mathbf{k}}$.

For a given vertex i of $\mathcal{G}_m$ with $k_i = k_{i1} + \cdots + k_{im}$, there are

$$\frac{k_i!}{k_{i1}! \cdots k_{im}!} = \binom{k_i}{k_{i1} \cdots k_{im}} \tag{25}$$

ways of departing from i. Clearly, (25) is the number of permutations with repetitions. Furthermore, $B_{\mathbf{k}}$ defined in (11) is the product of (25) for $i = 1, \ldots, m$. Let us define a *coalition* a set of m such permutations, one permutation per vertex, corresponding to a combination of the edges that depart from a vertex. There are $B_{\mathbf{k}}$ coalition.

Observe that for a given string, when scanning its symbols we trace an Eulerian path in $\mathcal{G}_m$. However, we are interested in an "inverse" problem: given an initial symbol $a \in \mathcal{A}$ and a matrix $\mathbf{k}$ satisfying the flow property (with a non zero weight for symbol a, $k_a > 0$), does a coalition of paths corresponds to a string x_1^n, that is, does it trace an Eulerian path. The problem is that such a trace may end prematurely at symbol $a \in \mathcal{A}$ (by exhausting all edges departing from a) without visiting all edges of $\mathcal{G}_m$ (i.e., the length of the traced string is shorter than n).[3] Let $\mathbf{k}'$ be the matrix composed of the remaining non-visited edges of the multigraph (the matrix $\mathbf{k}-\mathbf{k}'$ has been exhausted by the trace). Notice that matrix $\mathbf{k}'$ satisfies the flow property but the row and column corresponding to symbol a contain only zeros.

Given that $\mathbf{k}$ and $\mathbf{k}'$ are members of $\mathcal{F}_*$, let $N^a_{\mathbf{k},\mathbf{k}'}$ be the number of ways matrix $\mathbf{k}$ is transformed into another PO matrix $\mathbf{k}'$ when the Eulerian path starts with symbol a Notice that $k'_a = 0$. We have $N^a_{\mathbf{k},[0]} = N^a_{\mathbf{k}}$, but also the following

$$N^a_{\mathbf{k},\mathbf{k}'} = N^a_{\mathbf{k}-\mathbf{k}'} \times B_{\mathbf{k}'}, \qquad k'_a = 0.$$

[3]For example, in Figure 1 the following path 001010 of length six leaves edges 11 and 11 unvisited.

Summing over all matrices $\mathbf{k}'$ we obtain $\sum_{\mathbf{k}'} N^a_{\mathbf{k},\mathbf{k}'} = B_{\mathbf{k}}$, thus

$$B_{\mathbf{k}} = \sum_{\mathbf{k}',k'_a=0} N^a_{\mathbf{k}-\mathbf{k}'} \times B_{\mathbf{k}'}.$$

Multiplying by $\mathbf{z}^{\mathbf{k}}$ and summing now over all $\mathbf{z}^{\mathbf{k}}$ such that $k_a \neq 0$ it yields

$$\sum_{\mathbf{k}\in\mathcal{F}_*,k_a\neq 0} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \left(\sum_{\mathbf{k}} N^a_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\right).\times\left(\sum_{\mathbf{k}\in\mathcal{F}_*,k_a=0} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\right).$$

Denoting $N^a(\mathbf{z}) = \sum_{\mathbf{k}\in\mathcal{F}_*} N^a_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$, we finally arrive at

$$B_{\mathcal{A}}(\mathbf{z}) - B_{\mathcal{A}-\{a\}}(\mathbf{z}) = N^a(\mathbf{z})B_{\mathcal{A}-\{a\}}(\mathbf{z}).$$

We observe that for any generating functions $g(\mathbf{z})$ and $h(\mathbf{z})$ we have $\mathcal{F}(g\mathcal{F}h)(\mathbf{z}) = \mathcal{F}g(z)\mathcal{F}h(\mathbf{z})$. Consequently, $\mathcal{F}(\frac{1}{g})(\mathbf{z}) = \frac{1}{\mathcal{F}g(\mathbf{z})}$. Since $\mathcal{F}B(\mathbf{z}) = B_{\mathcal{A}}(\mathbf{z})$ and $\mathcal{F}B_{\mathcal{A}-\{a\}}(\mathbf{z}) = B_{\mathcal{A}-\{a\}}(\mathbf{z})$, for all $\mathbf{k}\in\mathcal{F}_*$ we finally arrive at

$$\begin{aligned}
[\mathbf{z}^{\mathbf{k}}]\frac{B_{\mathcal{A}}(\mathbf{z})}{B_{\mathcal{A}-\{a\}}(\mathbf{z})} &= [\mathbf{z}^{\mathbf{k}}]\frac{\mathcal{F}B(\mathbf{z})}{B_{\mathcal{A}-\{a\}}(\mathbf{z})} \\
&= [\mathbf{z}^{\mathbf{k}}]\mathcal{F}\left(\frac{B}{B_{\mathcal{A}-\{a\}}}\right)(\mathbf{z}) \\
&= [\mathbf{z}^{\mathbf{k}}]\frac{B(\mathbf{z})}{B_{\mathcal{A}-\{a\}}(\mathbf{z})},
\end{aligned}$$

which is the last step needed to complete the proof.

Knowing $N^a_{\mathbf{k}}$ we certainly can computed the frequency count $N_{\mathbf{k}}$ as

$$N_{\mathbf{k}} = [\mathbf{z}^{\mathbf{k}}]B(\mathbf{z})\sum_{a\in\mathcal{A}}(B_{\mathcal{A}-\{a\}}(\mathbf{z}))^{-1} = [\mathbf{z}^{\mathbf{k}}]B(\mathbf{z})\sum_{a\in\mathcal{A}}\det_{aa}(\mathbf{I}-\mathbf{z}).$$

3.2.2 Proof of Theorem 1(ii)

We establish now Theorem 1(ii), that is, we prove that *for $n \geq 1$ and $\mathbf{k}\in\mathcal{F}_n$, the frequency count $N^{ba}_{\mathbf{k}}$ is the coefficient of $\mathbf{z}^{\mathbf{k}}$ in* $\frac{B(\mathbf{z})z_{b,a}}{B_{\mathcal{A}-\{b\}}(\mathbf{z})} = z_{ba}B(\mathbf{z})\cdot\det_{bb}(\mathbf{I}-\mathbf{z})$.

The proof proceeds in the same way as in the previous theorem except that we have to consider a coalition with the first edge departing from b to a. We let $B^{ba}_{\mathbf{k}}$ be the number of such coalition. Observe that $B^{ba}_{\mathbf{k}} = B_{\mathbf{k}}\frac{k_{ba}}{k_b} = B_{\mathbf{k}-[\delta_{ba}]}$, where $[\delta_{ba}]$ is the matrix with all zeros except the ba-th element which is set to be one. Let $\mathbf{k}\in\mathcal{F}_*$. Then, using the same approach as before, we arrive at the following recurrence

$$B^{ba}_{\mathbf{k}} = \sum_{\mathbf{k}',k'_b=0} N^{ba}_{\mathbf{k}-\mathbf{k}'} \times B_{\mathbf{k}'}, \qquad k'_b = 0.$$

Computing the generating function we find

$$\sum_{\mathbf{k}\in\mathcal{F}_*,k_{ba}\neq 0} B^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \left(\sum_{\mathbf{k}} N^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\right)\times\left(\sum_{\mathbf{k}\in\mathcal{F}_*,k_b=0} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}\right).$$

In other words, $\sum_{\mathbf{k}\in\mathcal{F}_*,k_{ba}\neq 0} B^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = N^{ba}(\mathbf{z})B_{\mathcal{A}-\{a\}}(\mathbf{z})$, where $N^{ba}(\mathbf{z}) = \sum_{\mathbf{k}} N^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}}$. Using the fact that

$$\sum_{\mathbf{k}\in\mathcal{F}_*,k_{ba}\neq 0} B^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \mathcal{F}B^{ba}(\mathbf{z}),$$

where

$$B^{ba}(\mathbf{z}) \quad = \quad \sum_{\mathbf{k},k_{ba}>0} B^{ba}_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \sum_{\mathbf{k},k_{ba}>0} B_{\mathbf{k}-[\delta_{ba}]}\mathbf{z}^{\mathbf{k}} = B(\mathbf{z})z_{ba},$$

we complete the proof.

3.2.3 Proof of Theorem 1(iii)

Finally we establish Theorem 1(iii), that is, we prove that *for a given PO matrix* $\mathbf{k}$ *such that* $k_{ba} > 0$*, and* $k_{ij} = \Theta(n)$*,* $i, j \in \mathcal{A}$ *the following holds for large* n

$$N^{b,a}_{\mathbf{k}} = \frac{k_{ba}}{k_b} B_{\mathbf{k}} \cdot \det_{bb}(\mathbf{I} - \mathbf{k}^*) \left(1 + O\left(\frac{1}{n}\right)\right) \tag{26}$$

where $\mathbf{k}^*$ *is the matrix whose* ij*-th coefficient is* k_{ij}/k_i*, that is,* $\mathbf{k}^* = [k_{ij}/k_i]$.

From Cauchy's formula, (12) and Theorem 1(ii) we conclude that

$$B_{\mathbf{k}} \quad = \quad \left(\frac{1}{2\mathbf{i}\pi}\right)^{m^2} \oint \frac{B(\mathbf{z})}{\mathbf{z}^{\mathbf{k}+1}} d\mathbf{z}, \tag{27}$$

$$N^{ba}_{\mathbf{k}} \quad = \quad \left(\frac{1}{2\mathbf{i}\pi}\right)^{m^2} \oint \frac{B(\mathbf{z})z_{ba}\det(\mathbf{I}-\mathbf{z})}{\mathbf{z}^{\mathbf{k}+1}} d\mathbf{z}. \tag{28}$$

Recall that $B(\mathbf{z}) = \sum_{\mathbf{k}} B_{\mathbf{k}}\mathbf{z}^{\mathbf{k}} = \prod_i (1 - \sum_j z_{ij})^{-1}$. We make the change of variable $z_{ij} = \frac{k_{ij}}{k_i} e^{-\mathbf{i}t_{ij}/k_{ij}}$, where as before $k_i = \sum_j k_{ij}$. Observe that $\mathbf{z} = \mathbf{k}^* + O(1/n))$ where $\mathbf{k}^* = [k_{ij}/k_i]$, that is, it is a matrix with the ij-th coefficient equal to k_{ij}/k_i. More precisely,

$$\begin{aligned} 1 - \sum_j z_{ij} \quad &= \quad \sum_j \frac{k_{ij}}{k_i}(1 - e^{-\mathbf{i}t_{ij}/k_{ij}}) \\ &= \quad \frac{\mathbf{i}}{k_i}\left(\sum_j t_{ij}\right)\left(1 + O\left(\frac{1}{n}\right)\right), \qquad k_{ij} = O(n). \end{aligned}$$

Thus

$$\begin{aligned} B_{\mathbf{k}} \quad &= \quad (1 + O(n^{-1}))\prod_i \frac{\prod_j k_{ij}^{k_{ij}-1}}{k_i^{k_i-1}} \prod_{ij}\int_{-k_{ij}\pi}^{k_{ij}\pi} dt_{ij} \prod_i \left(\frac{1}{\sum_j t_{ij}}\right)\exp\left(\mathbf{i}\sum_j t_{ij}\right) \\ N^{ba}_{\mathbf{k}} \quad &= \quad (1 + O(n^{-1}))\prod_i \frac{\prod_j k_{ij}^{k_{ij}-1}}{k_i^{k_i-1}} \prod_{ij}\int_{-k_{ij}\pi}^{k_{ij}\pi} dt_{ij} \prod_i \left(\frac{1}{\sum_j t_{ij}}\right)\exp\left(\mathbf{i}\sum_j t_{ij}\right)\det(\mathbf{I}-\mathbf{z})z_{ba}. \end{aligned}$$

Since the function $\det(\mathbf{I}-\mathbf{z})z_{ba}$ is defined and bounded in a neighborhood of $\mathbf{k}^*$, we have $\det(\mathbf{I}-\mathbf{z})\mathbf{z}_{ba} = \det(\mathbf{I}-\mathbf{k}^*)k^*_{na}(1 + O(\frac{1}{n}))$. From this and (27)–(28) we conclude that

$$N^{ba}_{\mathbf{k}} = \left(1 + O\left(\frac{1}{n}\right)\right) B_{\mathbf{k}} k^*_{ba}\det(\mathbf{I}-\mathbf{k}^*).$$

This completes the proof.

3.3 Proof of Theorem 2

We finally prove our main result, namely Theorem 2. We start with estimating the size of $|\mathcal{F}_n|$, that is, the number of matrices $\mathbf{k}$ satisfying (9)–(10).

Lemma 3 *We have*

$$\frac{|\mathcal{F}_n|}{a(n)} = 1 + O(1/n),$$

where $a(n)$, defined above Lemma 2, is the volume of the simplex $\mathcal{K}(n)$.

Proof First, we give an estimate of $a(n)$. By setting $g(\mathbf{x}) = 1$ in Lemma 2 we find

$$\int_0^\infty a(x)e^{-tx}dx = \tilde{G}(t[1])$$

where $[1]$ is the matrix with all coefficients equal to 1. In order to estimate $a(x)$ we need the (multidimensional) Laplace of $g(\mathbf{x}) = 1$ which is

$$G(\mathbf{t}) = \int \exp(-\sum_{ij} t_{ij}x_{ij})d\mathbf{x} = \prod_{ij} \frac{1}{t_{ij}}.$$

Therefore, by (23) of Lemma 2 and the inverse Laplace we find

$$a(n) = \frac{1}{(2\mathbf{i}\pi)^{m+1}} \int_{c-\mathbf{i}\infty}^{c+\mathbf{i}\infty} dt \int_{c-\mathbf{i}\infty}^{c+\mathbf{i}\infty} d\theta_1 \cdots \int_{c-\mathbf{i}\infty}^{c+\mathbf{i}\infty} d\theta_m e^{nt} \prod_{ij} \frac{1}{t+\theta_i-\theta_j}$$

where $c > 0$. With the change of variable $(t', \theta'_1, \ldots, \theta'_m) = n(t, \theta_1, \ldots, \theta_m)$ we obtain

$$a(n) = \frac{n^{m^2-m-1}}{(2\mathbf{i}\pi)^{m+1}} \int dt' \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta'_1 \cdots \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta'_m e^{t'} \prod_{ij} \frac{1}{t'+\theta'_i-\theta'_j}. \tag{29}$$

Now, we turn to $|\mathcal{F}_n|$. We set $g(\mathbf{z}) = 1$ in Lemma 1 and define $F(z) = \sum_n |\mathcal{F}_n| z^n$. Observe that $F(z) = \mathcal{F}G(z[1])$ where $G(\mathbf{z}) = \sum_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} = \prod_{ij}(1-z_{ij})^{-1}$, and $z[1]$ is the matrix $\mathbf{z}$ with $\mathbf{z}_{ij} = z$ for $i, j \in \mathcal{A}$. By Lemma 1

$$\mathcal{F}G(\mathbf{z}) = \left(\frac{1}{2\mathbf{i}\pi}\right)^m \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_1 \cdots \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_m \prod_{ij}(1 - z_{ij}\exp(\theta_j - \theta_i))^{-1},$$

and therefore

$$F(z) = \left(\frac{1}{2\mathbf{i}\pi}\right)^m \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_1 \cdots \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_m \prod_{ij}(1 - z\exp(\theta_j - \theta_i))^{-1}.$$

Then by Cauchy's formula

$$\begin{aligned}
|\mathcal{F}_n| &= \frac{1}{2\mathbf{i}\pi} \oint \frac{dz}{z^{n+1}} F(z) \\
&= \left(\frac{1}{2\mathbf{i}\pi}\right)^{m+1} \oint \frac{dz}{z} \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_1 \cdots \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_m \prod_{ij}(1 - z\exp(\theta_j - \theta_i))^{-1} \frac{1}{z^n}.
\end{aligned}$$

With the change of variable $z = e^{-t}$ we find

$$|\mathcal{F}_n| = \left(\frac{1}{2\mathbf{i}\pi}\right)^{m+1} \int dt \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_1 \cdots \int_{-\mathbf{i}\pi}^{+\mathbf{i}\pi} d\theta_m \prod_{ij} (1-\exp(-t+\theta_j-\theta_i))^{-1} e^{nt}.$$

Let $(t', \theta'_1, \ldots, \theta'_m) = n(t, \theta_1, \ldots, \theta_m)$, then $1-\exp(-t-\theta_i+\theta_j) = \frac{1}{n}(t'+\theta'_i-\theta'_j)\left(1+O\left(\frac{1}{n}\right)\right)$, and finally we arrive at

$$\begin{aligned}
|\mathcal{F}_n| &= \frac{n^{m^2-m-1}}{(2\mathbf{i}\pi)^{m+1}}\left(1+O\left(\frac{1}{n}\right)\right) \int dt' e^{t'} \int_{-\mathbf{i}n\pi}^{+\mathbf{i}n\pi} d\theta'_1 \cdots \int_{-\mathbf{i}n\pi}^{+\mathbf{i}n\pi} d\theta'_m \prod_{ij} \frac{1}{t'+\theta'_i-\theta'_j} \\
&= \frac{n^{m^2-m-1}}{(2\mathbf{i}\pi)^{m+1}}\left(1+O\left(\frac{1}{n}\right)\right) \int dt' e^{t'} \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta'_1 \cdots \int_{-\mathbf{i}\infty}^{+\mathbf{i}\infty} d\theta'_m \prod_{ij} \frac{1}{t'+\theta'_i-\theta'_j} \\
&= a(n)\left(1+O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the last equality follows from (29). This completes the proof. ■

Now we are ready to prove Theorem 2 which is our main result. It suffices to calculate the partial redundancy $D_n^a(\mathcal{M}_1)$ restricted to all strings starting with a symbol a since $D_n(\mathcal{M}_1) = mD_n^a(\mathcal{M}_1)$. We have from (16)

$$D_n^a(\mathcal{M}_1) = \sum_{b\in\mathcal{A}} \sum_{\mathbf{k}\in\mathcal{F}_n, k_{ba}>0} \frac{N_{\mathbf{k}}^{ba}}{B_{\mathbf{k}}} B_{\mathbf{k}}(\mathbf{k}-[\delta_{ba}])^{\mathbf{k}-[\delta_{ba}]}(k_b-1)^{-k_b+1}\prod_{i\neq b}(k_i)^{-k_i} =$$

$$= (1+O(n^{-1}))\sum_{b\in\mathcal{A}} \sum_{\mathbf{k}\in\mathcal{F}_n, k_{ba}>0} \frac{k_{ba}}{k_a}\det_{bb}(\mathbf{I}-\mathbf{k}^*) B_{\mathbf{k}}(\mathbf{k}-[\delta_{ba}])^{\mathbf{k}-[\delta_{ba}]}(k_b-1)^{k_b-1}\prod_{i\neq b}(k_i)^{-k_i}.$$

Using Stirling's formula we obtain for $\mathbf{k}\in\mathcal{F}_n$ and $k_{ij} = \Theta(n)$

$$\frac{k_{ba}}{k_b} B_{\mathbf{k}}(\mathbf{k}-[\delta_{ba}])^{\mathbf{k}-[\delta_{ba}]}(k_b-1)^{-k_b+1}\prod_{i\neq b}(k_i)^{-k_i} = \prod_i \frac{\sqrt{2\pi k_i}}{\prod_j \sqrt{2\pi k_{ij}}}(1+O(1/n)),$$

and this yields

$$D_n^a(\mathcal{M}_1) = (1+O(1/n))\sum_{\mathbf{k}\in\mathcal{F}_n} F_m(\mathbf{k}^*)\prod_i \frac{\sqrt{2\pi k_i}}{\prod_j \sqrt{2\pi k_{ij}}},$$

where $F_m(\mathbf{x}) = \sum_{b\in\mathcal{A}} \det_{bb}(\mathbf{I}-\mathbf{x}^*)$ and $\mathbf{x}^*$ is the matrix whose (i,j) coefficient is x_{ij}/x_i, with $x_i = \sum_{j'} x_{ij'}$.

Using the Euler–Maclaurin summation formula (cf. [21]), we finally arrive at

$$D_n(\mathcal{M}_1) = \left(1+O\left(\frac{1}{n}\right)\right)\frac{|\mathcal{F}_n|}{a(n)}\int_{\mathcal{K}(n)} F_m(\mathbf{y})\prod_i \frac{\sqrt{2\pi\sum_j y_{ij}}}{\prod_j\sqrt{2\pi y_{ij}}}d\mathbf{y}. \tag{30}$$

Via trivial change of variable $\mathbf{y}' = \frac{1}{n}\mathbf{y}$, and since $F_m(\frac{1}{n}\mathbf{y}) = F_m(\mathbf{y})$ (indeed, $y_{ij}/y_i = y'_{ij}/y'_i$), we find

$$\int_{\mathcal{K}(n)} F_m(\mathbf{y})\prod_i \frac{\sqrt{2\pi\sum_j y_{ij}}}{\prod_j\sqrt{2\pi y_{ij}}}d\mathbf{y} = \left(\frac{n}{2\pi}\right)^{(m-1)m/2}\int_{\mathcal{K}(1)} F_m(\mathbf{y}')\prod_i \frac{\sqrt{\sum_j y'_{ij}}}{\prod_j\sqrt{y'_{ij}}}d\mathbf{y}'. \tag{31}$$

Since $|\mathcal{F}_n|/a(n) = 1 + O(1/n)$, we obtain our final result, that is,

$$D_n(\mathcal{M}_1) = \left(1 + O\left(\frac{1}{n}\right)\right)\left(\frac{n}{2\pi}\right)^{(m-1)m/2} A_m$$

for large n.

3.4 Redundancy of Markov Sources of Higher Order

We now sketch the proof of Theorem 3 for the maximal redundancy of universal codes for Markov sources of order r.

For Markov of order r, we define the PO matrix $\mathbf{k}$ as an $m^r \times m$ matrix whose $k_{w,j}$th coefficient ($w \in \mathcal{A}^r$) is the number of times the string w is followed by symbol $j \in \mathcal{A}$ in the string x_1^n. We can also view $\mathbf{k}$ as a $m^r \times m^r$ matrix indexed by $w, w' \in \mathcal{A}^r \times \mathcal{A}^r$ with the convention that nonzero elements of $\mathbf{k}$ are for $w' = w_2 \dots w_r j$, $j \in \mathcal{A}$, that is, when w' is constructed from w by deleting the first symbol and adding symbol $j \in \mathcal{A}$ at the end. Then

$$\sup_{P \in \mathcal{M}_r} P(x_1^n) = \prod_{w,j \in \mathcal{A}^{r+1}} \left(\frac{k_{w,j}}{k_w}\right)^{k_{w,j}}$$

where $k_w = \sum_j k_{w,j}$.

The main combinatorial result that we need is the enumeration of types, that is, how many strings of length n have type corresponding to the PO matrix $\mathbf{k}_{w,w'}$, $w, w' \in \mathcal{A}^r$ with w' defined above. As in the previous section, we focus on cyclic strings in which the last symbol is followed by the first. To enumerate cyclic strings of type $\mathbf{k}_{w,w'}$ we build a multigraph on m^r vertices with edges labeled by symbols from the alphabet $\mathcal{A}$. The number of Eulerian paths is equal to the number $\mathbf{N}_k$ of strings of type $\mathbf{k}$.

As in Section 3.2 we define $B_{\mathbf{k}}^r$ as the number of permutations with repetitions, that is,

$$B_{\mathbf{k}}^r = \prod_{w \in \mathcal{A}^r} \binom{k_w}{k_{w1} \cdots k_{wm}}.$$

Its generating function is

$$B^r(\mathbf{z}) = \prod_{w \in \mathcal{A}^r} (1 - \sum_{j \in \mathcal{A}} z_{w,j})^{-1}$$

while its $\mathcal{F}$ generating function of $B_{\mathbf{k}}^r$ is

$$\mathcal{F}B_r(\mathbf{z}) = (\det(\mathbf{I} - \mathbf{z}_r))^{-1},$$

where $\mathbf{z}_r$ is an $m^r \times m^r$ matrix whose (w, w') coefficient is equal to $z_{w,a}$ if there exist $a \in \mathcal{A}$ such that w' is a suffix of wa, otherwise the (w, w') coefficient is equal to 0 (as discussed above). Finally, we need to estimate $N_{\mathbf{k}}^{w,w'}$ the number of strings of type $\mathbf{k}$ starting with $ww' \in \mathcal{A}^{2r}$. As in Theorem 1(ii) we find that

$$N_{\mathbf{k}}^{w,w'} = [\mathbf{z}^{\mathbf{k}}] \left(B_r(\mathbf{z}) \det_{w,w}(\mathbf{I} - \mathbf{z}_r) \prod_{i=1}^{r} z_{(ww')_i^{i+r}} \right)$$

where $w_i^j = w_i w_{i+1} \dots w_j$ ($i \le j$). The rest follows the footsteps of our previous analysis and is omitted for brevity.

Acknowledgments

We thank Marcelo Weinberger (HPL, Palo Alto) for pointing to us the paper by Whittle, and Christian Krattenthaler (Vienna and Lyon) for showing us a connection between enumeration of spanning trees and Eulerian paths in graph.

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] L. Boza, Asymptotically Optimal Tests for Finite Markov Chains, *Ann. Math. Statistics*, 42, 1992-2007, 1971.

[5] R. Corless, G. Gonnet, D. Hare, D. Jeffrey and D. Knuth, On the Lambert W Function, *Adv. Computational Mathematics*, 5, 329–359, 1996.

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

[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, preprint; see also *Proc. LATIN 2002*, Springer LNCS 2286, 306-318, Cancun, Mexico, 2002.

[10] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM J. Disc. Methods*, 3, 216-240, 1990.

[11] J. Kieffer and E-H. Yang, Grammar-based Codes: A New Class of Universal Lossless Source Codes, *IEEE Trans. Information Theory*, 46, 737-754, 2000.

[12] R. Krichevsky and V. Trifimov, The Performance of Universal Coding, *IEEE Trans. Information Theory*, 27, 199–207, 1981.

[13] J. Rissanen, Complexity of Strings in the Class of Markov Sources, *IEEE Trans. Information Theory*, 30, 526–532, 1984.

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

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

[16] Y. Shtarkov, Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, 23, 175–186, 1987.

[17] Y. Shtarkov, T. Tjalkens and F.M. Willems, Multi-alphabet Universal Coding of Memoryless Sources, *Problems of Information Transmission*, 31, 114-127, 1995.

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

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

[20] W. Szpankowski, Asymptotic Redundancy of Huffman (and Other) Block Codes, *IEEE Trans. Information Theory*, 46, 2434-2443, 2000.

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

[22] V. K. Trofimov, Redundancy of Universal Coding of Arbitrary Markov Sources, *Probl. Inform. Trans.*, 10, 16–24, 1974 (Russian); 289–295, 1974 (English transl).

[23] Q. Xie, A. Barron, Minimax Redundancy for the Class of Memoryless Sources, *IEEE Trans. Information Theory*, 43, 647-657, 1997.

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

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

BIOGRAPHICAL SKETCHES

Wojciech Szpankowski received the M.S. degree and the Ph.D. degree in Electrical and Computer Engineering from Technical University of Gdańsk in 1976 and 1980, respectively. Currently, he is Professor of Computer Science at Purdue University. Before coming to Purdue, he was Assistant Professor at Technical University of Gdańsk, Poland, and in 1984 he held Visiting Assistant Professor position at the McGill University, Canada. During 1992/1993 he was Professeur Invité in the Institut National de Recherche en Informatique et en Automatique, France, in the Fall of 1999 he was Visiting Professor at Stanford University, and in June 2001 he was Professeur Invité at the Université de Versailles, France.

His research interests cover analytic algorithmics, information theory, bioinformatics, analytic combinatorics and random structures, pattern matching, discrete mathematics, performance evaluation, stability problems in distributed systems, and applied probability. He has published the book *Average Case Analysis of Algorithms on Sequences*, John Wiley & Sons, 2001. He wrote about 150 papers on these topics. Dr. Szpankowski has served as a guest editor for several journals: in 2002 he edited with M. Drmota a special issue for COMBINATORICS, PROBABILITY, & COMPUTING on analysis of algorithms, and currently he is editing together with J. Kieffer and E-H. Yang a special issue of the IEEE TRANSACTION ON INFORMATION THEORY on "Problems on Sequences: Information Theory & Computer Science Interface". He is on the editorial board of THEORETICAL COMPUTER SCIENCE and and FOUNDATION AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY. He also serves as the Managing Editor of DISCRETE MATHEMATICS AND THEORETICAL COMPUTER SCIENCE for "Analysis of Algorithms". Dr. Szpankowski chaired several workshops: in 1999 the *Information Theory and Networking Workshop*, Metsovo, Greece; in 2000 the *Sixth Seminar on Analysis of Algorithms*, Krynica Morska, Poland; and in 2003 the *NSF Workshop on Information Theory and Computer Science Interface*, Chicago. In June 2004 he will chair the *10th Seminar on Analysis of Algorithms*, Berkeley, CA. He is a recipient of the Humboldt Fellowship, and AFOSR, NSF, NIH and NATO research grants.

Philippe Jacquet is a research director in INRIA. He graduated from Ecole Polytechnique in 1981 and from Ecole Nationale des Mines in 1984. He received his Ph.D. degree from Paris Sud University in 1989 and his habilitation degree from Versailles University in 1998. He is currently the head of HIPERCOM project that is devoted to high performance communications. As an expert in telecommunications and information technology, he participated in several standardization committees such as ETSI, IEEE and IETF. His research interests cover information theory, probability theory, quantum telecommunication, evaluation of performance and algorithm design for telecommunication, wireless and ad hoc networking.

Philippe Jacquet is author of numerous papers that have appeared in international journals. In 1999 he co-chaired the *Information Theory and Networking Workshop*, Metsovo, Greece.