



IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Average Redundancy of the Shannon Code for Markov Sources

**Neri Merhav and
Wojciech Szpankowski**

CCIT Report #811
August 2012

■ ■ ■ ■ Electronics
■ ■ ■ ■ Computers
■ ■ ■ ■ Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL



Average Redundancy of the Shannon Code for Markov Sources

Neri Merhav*

Wojciech Szpankowski†

August 6, 2012

Abstract

It is known that for memoryless sources, the average and maximal redundancy of fixed-to-variable length codes, such as the Shannon and Huffman codes, exhibit two modes of behavior for long blocks. It either converges to a limit or it has an oscillatory pattern, depending on the irrationality or rationality, respectively, of certain parameters that depend on the source. In this paper, we extend these findings, concerning the Shannon code, to the case of a Markov source, which is considerably more involved. While this dichotomy, of convergent vs. oscillatory behavior, is well known in other contexts (including renewal theory, ergodic theory, local limit theorems and large deviations of discrete distributions), in information theory (e.g., in redundancy analysis) it was recognized relatively recently. To the best of our knowledge, no results of this type were reported thus far for Markov sources. We provide a precise characterization of the convergent vs. oscillatory behavior of the Shannon code redundancy for a class of irreducible, periodic and aperiodic, Markov sources. These findings are obtained by analytic methods, such as Fourier/Fejér series analysis and spectral analysis of matrices.

Index Terms: Shannon code, average redundancy, Fourier series, uniform convergence, spectral analysis, analytic information theory.

*N. Merhav is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel. E-mail: merhav@ee.technion.ac.il. N. Merhav's work was supported by the Israeli Science Foundation (ISF), grant no. 208/08.

†W. Szpankowski is with the Department of Computer Science, Purdue University, W. Lafayette, IN 47907, U.S.A. E-mail: spa@cs.purdue.edu. He is also a Visiting Professor at ETI, Gdańsk University of Technology, Poland. W. Szpankowski's work was supported in part by the NSF Science and Technology Center for Science of Information Grant CCF-0939370, NSF Grant CCF-0830140, AFOSR Grant FA8655-11-1-3076, and NSA Grant H98230-11-1-0141.

1 Introduction

Recent years have witnessed a resurgence of interest in redundancy rates of lossless coding, see, e.g., [1], [3], [6], [10], [13], [14], [15], [16], [17], [18]. In particular, in [18] Szpankowski derived asymptotic expressions of the (unnormalized) average redundancy R_n , as a function of the block length n , for the Shannon code, the Huffman code, and other codes, focusing primarily on the binary memoryless source (BSS), parametrized by p – the probability of ‘1’. A rather interesting behavior of R_n was revealed in [18], especially in the cases of the Shannon code and the Huffman code: When $\alpha \triangleq \log_2[(1-p)/p]$ is irrational, then R_n converges to a constant (which is $1/2$ for the Shannon code), as $n \rightarrow \infty$. On the other hand, when α is rational, R_n has a non-vanishing oscillatory term whose fundamental frequency and amplitude depend on the source statistics in an explicit manner.

More precisely, confining the discussion to the Shannon code, in [18] the average unnormalized redundancy

$$R_n = \mathbf{E} \{ \lceil -\log_2 P(X_1, \dots, X_n) \rceil + \log_2 P(X_1, \dots, X_n) \}, \quad (1)$$

was analyzed for large n , assuming that the source P , that governs the data to be compressed, X_1, X_2, \dots , is a BSS. A straightforward extension (see also [14]) of the Shannon-code redundancy result of [18], to a general r -ary alphabet memoryless source, with letter probabilities p_1, \dots, p_r , yields the following expression:

$$R_n = \begin{cases} \frac{1}{2} + \frac{1}{M} \left(\frac{1}{2} - \langle \beta M n \rangle \right) + o(1) & \text{all } \{\alpha_j\} \text{ are rational} \\ \frac{1}{2} + o(1) & \text{otherwise} \end{cases} \quad (2)$$

where $\beta \triangleq -\log p_1$, $\alpha_j = \log p_j/p_1$, $j = 2, 3, \dots, r$, $\langle u \rangle$ is the fractional part of a real number u (i.e., $\langle u \rangle = u - \lfloor u \rfloor$), and M is the smallest common multiple of all denominators of the rational numbers $\{\alpha_j\}$ when presented as ratios between two relatively prime integers. This erratic behavior, where R_n is either convergent (and then the limit is always $1/2$) or oscillatory, depending on the rationality of $\{\alpha_j\}$, was related in [14] to wave diffraction patterns of scattering from partially disordered media, where the existence/non-existence of Bragg peaks depends on the rationality/irrationality of certain optical distance ratios.

Our goal in this paper is to extend the scope of this analysis to irreducible Markov sources and to evaluate precisely (for large n) the average redundancy of the Shannon code for a finite alphabet, first order Markov source with given transition probabilities. In doing

so, we also provide a more complete analysis than in [14] and [18]. As will be seen, this extension to the Markov case appears rather non-trivial, both from the viewpoint of the conditions for oscillatory behavior and from the aspect of the asymptotic expression of R_n in the oscillatory mode. These depend strongly on the dominant eigenvalues and on the detailed structure of the matrix of transition probabilities. For example, in contrast to the memoryless case, where there is only one oscillatory term, when it comes to the Markov case, in the oscillatory mode there are, in general, contributions from multiple oscillatory terms, and in the convergent mode, R_n may converge to a constant other than $1/2$ (see Example 2 below). Moreover, it turns out that the behavior of the redundancy depends quite strongly on important dynamical properties of the Markov chain, such as reducibility/irreducibility and periodicity/apperiodicity.

We begin our study (Sections 2 and 3) from the relatively simple case where all single-step state transitions have positive probability. Our main result in Section 2, Theorem 1, is then an extension of formula (2) to the Markov case with strictly positive state transition probabilities. To give the reader a general idea of this theorem, an informal description of it can be stated as follows: Rather than the parameters $\{\alpha_j\}$ of the memoryless case, we now define a matrix $\{\alpha_{jk}\}_{j,k=1}^r$ of log-ratios of certain transition probabilities (the exact definition will be provided in the sequel). If at least one of these parameters is irrational, then similarly as in the memoryless case, $R_n = \frac{1}{2} + o(1)$. If, on the other hand, all these parameters are rational, then as in the memoryless case, let M be their smallest common denominator. In this case, $R_n = \Omega_n + o(1)$, for “most large values” of n (a term that will be defined precisely in the sequel), where Ω_n is a linear combination of certain functions of n , for which we have an explicit formula in terms of the source parameters. These functions oscillate as n varies, with amplitude $1/M$ and a fundamental frequency that depends on the source parameters.

In Section 4, we relax the strict positivity assumption, but still assume the Markov chain to be irreducible. Under this assumption, we first assume that the chain is also aperiodic, and then further extend the scope to allow periodicity. In these cases, the extension of eq. (2) is still available, though it is somewhat less explicit (than in the positive transition matrix case) in the sense that it depends on certain parameters of the source, for which we have no closed-form expressions, but which can be found by numerical procedures. It is also demonstrated (in Example 2) that the irreducibility assumption is essential, since the

above described two-mode behavior ceases to exist when this assumption is dropped.

We should point out that minimax redundancy and regret for the *class* of Markov sources were studied in the past – see, e.g., [10], [15]. Interestingly enough, the minimax regret for memoryless and Markov sources does not exhibit the two-mode behavior of either convergent or oscillatory mode [3]. This dichotomy, of convergent vs. oscillatory behavior, with dependence on rationality/irrationality of certain parameters, is a well recognized phenomenon in mathematics and physics, ranging across a large variety of areas, including renewal theory, ergodic theory [7], local limit theorems and large deviations for discrete distributions [2], [4]. This phenomenon, however, was observed in information theory only relatively recently [7], [18]. On the other hand, the oscillatory phenomenon for discrete random structures is a well known fact in analysis of algorithms [5], [19], and also in information theory [3], [13], [19].

2 Formulation and Results for Positive Transition Matrices

In this section, we first establish notation conventions and spell out our assumptions. Then, we present our main result for the case of a positive transition probability matrix (Theorem 1), discuss it, and provide an example for its use.

Throughout this paper, we adopt the customary notation conventions in the information theory literature: Random variables will be denoted by capital letters (e.g., X), specific values they may take will be denoted by the corresponding lower-case letters (e.g., x), and their alphabets will be denoted by the corresponding calligraphic letters (e.g., \mathcal{X}). Random vectors of length n (e.g., (X_1, X_2, \dots, X_n)) will be denoted by capital letters superscripted by n (e.g., X^n), and specific values of these vectors (e.g., (x_1, x_2, \dots, x_n)) will be denoted by lower-case letters superscripted by n (e.g., x^n). Finally, the set of vectors of length n , with components taking on values in \mathcal{X} , will be denoted by \mathcal{X}^n . Logarithms will always be understood to be taken w.r.t. the base 2. The function $\mathcal{I}(\cdot)$ will denote the indicator function, that is, for a given statement E , $\mathcal{I}(E) = 1$ if E is true, and $\mathcal{I}(E) = 0$ if E is false.

Consider a source sequence $X_1, X_2, \dots, X_t \in \mathcal{X} = \{1, 2, \dots, r\}$ (r – positive integer), $t = 1, 2, \dots$, governed by a first-order Markov chain with a given matrix P of state-transition probabilities $\{p(j|k)\}_{j,k=1}^r$. The initial state probabilities will be denoted by p_k , $k = 1, 2, \dots, r$. The stationary state probabilities will be denoted by π_k , $k = 1, 2, \dots, r$. Thus, the probability of a given source string $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$, under the given

Markov source, is

$$\mu(x^n) = p_{x_1} \prod_{t=2}^n p(x_t | x_{t-1}). \quad (3)$$

The average unnormalized redundancy of the Shannon code is defined as

$$R_n \triangleq \mathbf{E}\{\lceil -\log \mu(X^n) \rceil + \log \mu(X^n)\}, \quad (4)$$

where here and throughout the sequel, $\mathbf{E}\{\cdot\}$ denotes the expectation operator w.r.t. the underlying Markov source μ just defined.

As mentioned in the Introduction, in this paper, we assume that P is irreducible. We remind the reader that an *irreducible* Markov chain is one where there is positive probability to pass from every state $j \in \mathcal{X}$ to every state $k \in \mathcal{X}$ within a finite number of steps, namely, for every j and k , there exists a positive integer l such that the (k, j) -th element of P^l is strictly positive. Another important concept we will need is periodicity. The *period* d_j of a state j is the greatest common divisor of all integers n for which $\Pr\{X_n = j | X_0 = j\} > 0$. A state is called periodic if $d_j > 1$ and aperiodic if $d_j = 1$. Since all states of an irreducible Markov chain are in the same class of communicating states, then d_j is the same for all states, and hence will be denoted collectively by d . An irreducible Markov chain is then called periodic if $d > 1$ and aperiodic if $d = 1$. The case where all entries of P are positive, henceforth referred to as the case of a *positive matrix* P , is obviously a case of an irreducible, aperiodic Markov chain. However, the positivity of P is not a necessary condition for irreducibility and aperiodicity of a Markov chain. Throughout the remaining part of this section, as well as in Section 3, we assume that all entries of P are strictly positive.

Our main result in this section is the following (the proof appears in Section 3).

Theorem 1 *Consider the Shannon code of block length n for a Markov source μ with a given vector $\mathbf{p} = (p_1, \dots, p_r)$ of initial state probabilities and a positive state transition matrix P . Define*

$$\alpha_{jk} = \log \left[\frac{p(j|1)p(j|j)}{p(k|1)p(j|k)} \right], \quad j, k \in \{1, 2, \dots, r\}. \quad (5)$$

Then, the redundancy R_n is characterized as follows:

(a) If not all $\{\alpha_{jk}\}$ are rational, then

$$R_n = \frac{1}{2} + o(1). \quad (6)$$

(b) If all $\{\alpha_{jk}\}$ are rational, then for every $j, k \in \{1, \dots, r\}$, let

$$\zeta_{jk}(n) = M[-(n-1)\log p(1|1) + \log p(j|1) - \log p(k|1) - \log p_j], \quad (7)$$

and

$$\Omega_n = \frac{1}{2} \left(1 - \frac{1}{M}\right) + \frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \varrho[\zeta_{jk}(n)], \quad (8)$$

where $\varrho(u) \triangleq [u] - u$ and M is the smallest common integer multiple of the denominators of $\{\alpha_{jk}\}$, when each one of these numbers is represented as a ratio between two relatively prime integers. Then, there exists a positive sequence $\xi_n \rightarrow 0$, which depends only the source parameters, such that R_n is upper bounded and lower bounded as follows:

$$R_n \leq \Omega_n + \frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\} + o(1). \quad (9)$$

$$R_n \geq \Omega_n - \frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\} - o(1). \quad (10)$$

As a technical comment, it should be pointed out that the choice of the index 1 in the conditioning of $p(j|1)$ and $p(k|1)$, that appear in the definition of α_{jk} and in (7), is completely arbitrary. One may choose any other index in $\{1, 2, \dots, r\}$, as long as it is the same index in both places in the expression of α_{jk} , as well as in the second and third terms in the square brackets of (7). Also, $p(1|1)$ in (7) can be replaced independently by $p(l|l)$ for any $l \in \{1, 2, \dots, r\}$.

Discussion. Theorem 1 tells us that, similarly as in the memoryless case, in the positive matrix case, R_n has two modes of behavior. In the convergent mode, which happens when at least one α_{jk} is irrational, $R_n \rightarrow 1/2$. In the oscillatory mode, which happens when all $\{\alpha_{jk}\}$ are rational, R_n oscillates and it asymptotically coincides with Ω_n for *most large values*¹ of n , provided that $\log p(1|1)$ is irrational. This follows from the following consideration: If $\log p(1|1)$ is irrational, then by Weyl's equidistribution theorem [12], the sequences $\{\zeta_{jk}(n)\}_{n \geq 1}$ are uniformly distributed modulo 1, i.e., they fill the unit interval mod 1 with a uniform density as n exhausts the positive integers. Thus, for every fixed ξ , $\varrho[\zeta_{jk}(n)] \notin (\xi, 1 - \xi)$ for a fraction 2ξ of the values of n . This means that for $\xi_n \rightarrow 0$,

¹The statement " R_n asymptotically coincides with Ω_n for most large values of n " means that for every $\epsilon > 0$, the fraction of values of n , within the range $\{1, \dots, N\}$, for which $|R_n - \Omega_n| > \epsilon$, tends to zero as $N \rightarrow \infty$.

the terms $\mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\}$ vanish for most large values of n , and then the lower bound and the upper bound on R_n asymptotically coincide with Ω_n . If, on the other hand, $\log p(1|1)$ is rational, then $\varrho[\zeta_{jk}(n)]$ are periodic sequences. If for none of the values n in a period, $\varrho[\zeta_{jk}(n)] = 0$, then beyond a certain value of n , ξ_n is smaller than the minimum value of $\varrho[\zeta_{jk}(n)]$ along the period and $1 - \xi_n$ is larger than the maximum, and so, $\mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\}$ all vanish for *all* large n . The expression

$$\frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\},$$

which generates the gap between the upper bound and the lower bound on R_n , can be interpreted as an asymptotic approximation of the probability that $-\log \mu(X^n)$ falls in the vicinity (within distance $O(\xi_n)$) of an integer. For example, when the source is purely dyadic ($M = 1$), then $-\log \mu(X^n)$ is integer with probability 1, and indeed, the expression in the last display is equal to 1. In this case, Theorem 1 is useless, but it is also redundant, because in this case, we clearly know that R_n vanishes. The reason for this “uncertainty” around integer values of $-\log \mu(X^n)$ is that these are the discontinuity points of the function $\varrho[-\log \mu(X^n)]$, and in the proof of Theorem 1, the function ϱ is expanded as a series of trigonometric polynomials whose convergence is problematic in the neighborhood of discontinuities. Thus, we believe that the uncertainty in the characterization of R_n around these points should be attributed more to the limitations of the analysis methods than to the real behavior of R_n . In other words, we conjecture that, in fact, $R_n = \Omega_n + o(1)$ for *all* large n , and not just for most large values of n . It should be pointed out that these issues were admittedly overlooked in [14] and [18] (beyond the cases of a purely dyadic source, which was ruled out in the first place). The essential results therein are nonetheless re-confirmed here as a special case, upon carrying out a more rigorous analysis.

The expression of the oscillatory case, Ω_n , is not quite intuitive at first glance, therefore, in this paragraph, we make an attempt to give some quick insight, which captures the essence of the main points. The arguments here are informal and non-rigorous (the rigorous proof is in Section 3). The Fourier series expansion of the periodic function ϱ is given by

$$\varrho(u) = \frac{1}{2} + \sum_{m \neq 0} a_m e^{2\pi i m u} \quad (11)$$

and the important fact about the coefficients is that they are inversely proportional to m , so that for every two integers k and m , $a_{m \cdot k} = a_m/k$. Now, when computing $R_n =$

$\mathbf{E}\{\varrho[-\log \mu(X^n)]\}$, let us take the liberty of exchanging the order between the expectation and the summation, i.e.,

$$R_n = \frac{1}{2} + \sum_{m \neq 0} a_m \mathbf{E}\{e^{-2\pi i m \log \mu(X^n)}\}. \quad (12)$$

It turns out that under the conditions of the oscillatory mode, $\mathbf{E}\{e^{-2\pi i m \log \mu(X^n)}\}$ tends to zero as $n \rightarrow \infty$ for all m , except for multiples² of M , namely, $m = \ell M$, $\ell = \pm 1, \pm 2, \dots$. Thus, for large n , we have

$$\begin{aligned} R_n &\approx \frac{1}{2} + \sum_{\ell \neq 0} a_{\ell M} \mathbf{E}\{e^{-2\pi i \ell M \log \mu(X^n)}\} \\ &= \frac{1}{2} + \frac{1}{M} \sum_{\ell \neq 0} a_{\ell} \mathbf{E}\{e^{-2\pi i \ell M \log \mu(X^n)}\} \\ &= \frac{1}{2} + \frac{1}{M} \left\{ \mathbf{E}\varrho[-M \log \mu(X^n)] - \frac{1}{2} \right\} \\ &= \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \mathbf{E}\varrho[-M \log \mu(X^n)]. \end{aligned} \quad (13)$$

Now, consider the set of all $\{x^n\}$ that begin from state $x_1 = j$ and end at state $x_n = k$. Their total probability is about $p_j \pi_k$ for large n since X_n is almost independent of X_1 . It turns out that all these sequences have exactly the same value of $\varrho[-M \log \mu(x^n)]$, which is exactly $\varrho[\zeta_{jk}(n)]$ (or, in other words, $\varrho[-M \log \mu(x^n)] = \varrho[\zeta_{x_1 x_n}(n)]$ independently of x_2, \dots, x_{n-1}) and this explains the expression of Ω_n . The reason for this property of $\varrho[-M \log \mu(x^n)]$ is the rationality conditions $\langle M \cdot \alpha_{uv} \rangle = 0$, $u, v \in \{1, 2, \dots, r\}$, which imply that $\langle M \log p(x_t | x_{t-1}) \rangle = \langle M \log[p(x_t | 1)p(1 | 1)/p(x_{t-1} | 1)] \rangle$, and so,

$$\begin{aligned} \langle -M \log \mu(x^n) \rangle &= \langle -M \log p_j \rangle + \sum_{t=2}^n \langle -M \log p(x_t | x_{t-1}) \rangle \mod 1 \\ &= \langle -M \log p_j \rangle + \sum_{t=2}^n \langle -M \log[p(x_t | 1)p(1 | 1)/p(x_{t-1} | 1)] \rangle \mod 1 \end{aligned} \quad (14)$$

which, thanks to the telescopic summation, is easily seen to coincide with the fractional part of $\zeta_{jk}(n)$, and of course, $\varrho[\zeta_{jk}(n)]$ depends on $\zeta_{jk}(n)$ only via its fractional part.

Consider next the following example for using Theorem 1.

Example 1. Consider a Markov source for which the rows of P are all permutations of the first row, which is $\mathbf{p} = (p_1, \dots, p_r)$. Now, assuming that $\alpha_j \triangleq \log(p_1/p_j)$ are all rational, let M be the least common multiple of their denominators (i.e., the common denominator)

²The convergent mode can be treated as a special case of this statement with $M = \infty$.

when each one of them is expressed as a ratio between two relatively prime integers. Then,

$$\begin{aligned}
\varrho[\zeta_{jk}(n)] &= \varrho[-M(n-1)\log p(1|1) + M\log p(j|1) - M\log p(k|1) - M\log p_j] \\
&= \varrho[-M(n-1)\log p_1 + M\log p_j - M\log p_k - M\log p_j] \\
&= \varrho[-M(n-1)\log p_1 - M\log p_k] \\
&= \varrho(-Mn\log p_1 + M\log p_1 - M\log p_k) \\
&= \varrho(-Mn\log p_1),
\end{aligned} \tag{15}$$

where in the last step, we have used the fact that $(M\log p_1 - M\log p_k)$ is integer and that ϱ is a periodic function with period 1. Thus, with the exception of the minority of ‘problematic’ values of n , we have

$$\begin{aligned}
R_n &= \frac{1}{2} \left(1 - \frac{1}{M}\right) + \frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \varrho[\zeta_{jk}(n)] + o(1) \\
&= \frac{1}{2} \left(1 - \frac{1}{M}\right) + \frac{1}{M} \sum_{j=1}^r \sum_{k=1}^r p_j \pi_k \varrho(-nM\log p_1) + o(1) \\
&= \frac{1}{2} \left(1 - \frac{1}{M}\right) + \frac{1}{M} \varrho(-nM\log p_1) + o(1).
\end{aligned} \tag{16}$$

If not all α_j are rational, then $R_n \rightarrow 1/2$, as predicted by Theorem 1. To see why the conditions of Theorem 1 lead to the rationality condition herein, let us denote $u_{jk} = \langle m \log[p(j|1)/p(k|1)] \rangle$, and $v_{jk} = \langle m \log[p(j|j)/p(j|k)] \rangle$. Then, the conditions of Theorem 1 mean that $u_{jk} + v_{jk} = 0$ and for all pairs j and k . Therefore, the number of constraints here is of the order of r^2 , whereas the number of degrees of freedom that generate these variables, in this example, is $r - 1$, i.e., the variables $\langle m \log(p_1/p_j) \rangle$, $j = 2, 3, \dots, r$. Thus, we can think of this as an overdetermined set of homogeneous linear equations whose only solution is zero, meaning that $\langle m \log(p_1/p_j) \rangle$, $j = 2, 3, \dots, r$, all vanish. Note that the memoryless source is a special case of this example, where the rows of P are all identical to the first row, (p_1, \dots, p_r) . Indeed, eq. (16) coincides with the expression of the memoryless case (see [14], [18] and the Introduction of this paper).

3 Proof of Theorem 1

3.1 Introductory Comments

The main idea behind the analysis of $R_n = \mathbf{E}\{\varrho[-\log \mu(X^n)]\}$ is to approximate the periodic function $\varrho(\cdot)$ by a sequence of trigonometric polynomials, and then to commute the

expectation with the summation and analyze the various terms of the series. For these commutations to be legitimate, a sufficient condition is that the convergence would be uniform, but unfortunately, it cannot be uniform since the function ϱ is discontinuous. An alternative route that we take is to sandwich ϱ between two continuous periodic functions, ϱ_θ^- and ϱ_θ^+ , both with period 1, and both indexed by some parameter θ , which when tends to zero, the bounds become tighter and tighter. Fejér's theorem (see, e.g., [20]), which is the trigonometric version of the Weierstrass theorem, provides a concrete sequence of trigonometric polynomials, which converges uniformly to any given periodic function which is continuous. The program of the proof is to apply Fejér's theorem to ϱ_θ^- , and ϱ_θ^+ , and use them to obtain sandwich bounds on R_n .

3.2 Preliminaries of the Proof

Define the function ϱ_θ^- as

$$\varrho_\theta^-(u) = \begin{cases} \frac{1-\theta}{\theta} \cdot \langle u \rangle & 0 \leq \langle u \rangle < \theta \\ 1 - \langle u \rangle & \theta \leq \langle u \rangle < 1 \end{cases} \quad (17)$$

and

$$\varrho_\theta^+(u) = \varrho_\theta^-(u) + \Delta_\theta(u) \quad (18)$$

where

$$\Delta_\theta(u) = \begin{cases} 1 - \frac{\langle u \rangle}{\theta} & 0 \leq \langle u \rangle < \theta \\ 0 & \theta \leq \langle u \rangle < 1 - \theta \\ \frac{1}{\theta}(\langle u \rangle + \theta - 1) & 1 - \theta \leq \langle u \rangle < 1 \end{cases} \quad (19)$$

Obviously, $\varrho_\theta^-(u)$, and $\varrho_\theta^+(u)$ are continuous, periodic functions, with period 1, and $\varrho_\theta^-(u) \leq \varrho(u) \leq \varrho_\theta^+(u)$ for every u . Now, ϱ_θ^- and Δ_θ have the following Fourier representations:

$$\varrho_\theta^-(u) = \frac{1}{2} + \sum_{m \neq 0} a_m(\theta) e^{2\pi i m u}; \quad a_m(\theta) = \frac{1 - e^{-2\pi i m \theta}}{(2\pi i m)^2 \theta} \quad (20)$$

and

$$\Delta_\theta(u) = \theta + \sum_{m \neq 0} b_m(\theta) e^{2\pi i m u}; \quad b_m(\theta) = \frac{1 - \cos(2\pi m \theta)}{2\theta \pi^2 m^2}. \quad (21)$$

Note that for any given integers k and ℓ ,

$$a_{\ell \cdot k}(\theta) = \frac{a_\ell(k\theta)}{k} \quad (22)$$

and similarly

$$b_{\ell \cdot k}(\theta) = \frac{b_\ell(k\theta)}{k}. \quad (23)$$

These identities will be important later on, in order to return from the series expansions back to the original functions. The N -the order Féjer approximations are given by

$$\{\varrho_{\theta}^{-}(u)\}_N \triangleq \frac{1}{2} + \sum_{|m|=1}^N a_m(\theta) \cdot \left(1 - \frac{|m|}{N+1}\right) e^{2\pi i m u} \quad (24)$$

and

$$\{\Delta_{\theta}(u)\}_N \triangleq \theta + \sum_{|m|=1}^N b_m(\theta) \cdot \left(1 - \frac{|m|}{N+1}\right) e^{2\pi i m u}. \quad (25)$$

According to Fejér's theorem, as $N \rightarrow \infty$, these functions converge uniformly to $\varrho_{\theta}^{-}(u)$ and $\Delta_{\theta}(u)$, respectively. However, it should be kept in mind that in order to guarantee that the absolute error would be uniformly within less than a given ϵ (for all three functions ρ_{θ}^{+} , ρ_{θ}^{-} , and Δ_{θ}), the integer N should be at least as large as some $N_0(\epsilon, \theta)$ (or N_0 for shorthand notation), which grows both as ϵ decreases and as θ decreases. In particular, following the proof of Fejér's theorem [20, p. 6] (see also Appendix herein), it is readily seen that for all three functions, ρ_{θ}^{+} , ρ_{θ}^{-} , and Δ_{θ} ,

$$\epsilon_0(N, \theta) \triangleq \inf_{0 < \delta < 1/2} \left[\frac{\delta}{\theta} + \frac{1}{N \sin^2(\pi \delta)} \right] \quad (26)$$

is an upper bound on the maximum approximation error when N terms of the Fejér series are used. Thus, $N_0(\epsilon, \theta)$ can be defined as the smallest integer N such that $\epsilon_0(N, \theta) \leq \epsilon$. Obviously, by definition

$$\epsilon_0[N_0(\epsilon, \theta), \theta] \leq \epsilon. \quad (27)$$

We will make use of this simple inequality later on.

3.3 General Lower and Upper Bounds on R_n

We proceed with some general lower and upper bounds on R_n . As for the lower bound, we have

$$\begin{aligned} R_n &= \mathbf{E} \{ \varrho(-\log \mu(X^n)) \} \\ &\geq \mathbf{E} \{ \varrho_{\theta}^{-}(-\log \mu(X^n)) \} \\ &\geq \mathbf{E} \left\{ \frac{1}{2} + \sum_{|m|=1}^{N_0} a_m(\theta) \cdot \left(1 - \frac{|m|}{N_0+1}\right) e^{-2\pi i m \log \mu(X^n)} - \epsilon \right\} \\ &= \frac{1}{2} + \sum_{|m|=1}^{N_0} a_m(\theta) \cdot \left(1 - \frac{|m|}{N_0+1}\right) \mathbf{E} \{ e^{-2\pi i m \log \mu(X^n)} \} - \epsilon. \end{aligned} \quad (28)$$

Now, clearly

$$\mathbf{E} \left\{ e^{-2\pi i m \log \mu(X^n)} \right\} = \sum_{\mathbf{x} \in \mathcal{X}^n} \prod_{t=1}^n [p(x_t | x_{t-1}) \exp \{-2\pi i m \log p(x_t | x_{t-1})\}]. \quad (29)$$

Define the $r \times r$ complex matrix A_m whose entries are

$$a_{jk}(m) = p(k|j) \exp[-2\pi i m \log p(k|j)], \quad j, k = 1, \dots, r. \quad (30)$$

Also define the r -dimensional column vectors

$$\mathbf{c}_m = (p_1 \exp[-2\pi i m \log p_1], \dots, p_r \exp[-2\pi i m \log p_r])^T, \quad (31)$$

and $\mathbf{1} = (1, 1, \dots, 1)^T$, where the superscript T denotes vector/matrix transposition. Then, it follows that

$$\mathbf{E} \left\{ e^{-2\pi i m \log \mu(X^n)} \right\} = \mathbf{c}_m^T A_m^{n-1} \mathbf{1}. \quad (32)$$

Let $\mathbf{l}_{j,m}$ and $\mathbf{r}_{j,m}$ be, respectively, the left eigenvector and the right eigenvector pertaining to the eigenvalue $\lambda_{j,m}$ ($j = 1, 2, \dots, r$) of the matrix A_m . Here, we index the eigenvalues of A_m according to a non-increasing order of their modulus, that is,

$$|\lambda_{1,m}| \geq |\lambda_{2,m}| \geq \dots \geq |\lambda_{r,m}|. \quad (33)$$

Since P is a stochastic matrix (so, its maximum modulus eigenvalue is 1) and its elements are the absolute values of the corresponding elements of A_m , it follows from [8, Theorem 8.4.5] (see also Lemma 1 in Subsection 3.4) that $|\lambda_{1,m}| \leq 1$ (and hence $|\lambda_{j,m}| \leq 1$ for all $j = 1, 2, \dots, r$). Also, the sets of left- and right eigenvectors form a bi-orthogonal system, i.e., $\mathbf{l}_{j,m}^T \mathbf{r}_{k,m} = 0$, $j, k = 1, 2, \dots, r$, $j \neq k$. We scale these vectors such that $\mathbf{l}_{j,m}^T \mathbf{r}_{j,m} = 1$ for all $j = 1, 2, \dots, r$. Then by the spectral representation of matrices [8], we have

$$A_m^{n-1} \mathbf{1} = \sum_{j=1}^r \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{r}_{j,m}, \quad (34)$$

and so,

$$\mathbf{c}_m^T A_m^{n-1} \mathbf{1} = \sum_{j=1}^r \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m}. \quad (35)$$

On substituting this back into the lower bound on R_n , we obtain:

$$R_n \geq \frac{1}{2} + \sum_{|m|=1}^{N_0} a_m(\theta) \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \cdot \sum_{j=1}^r \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m} - \epsilon. \quad (36)$$

In a similar manner, we obtain the following upper bound

$$\begin{aligned}
R_n &= \mathbf{E} \{ \varrho(-\log \mu(X^n)) \} \\
&\leq \mathbf{E} \{ \varrho_\theta^+(-\log \mu(X^n)) \} \\
&= \mathbf{E} \{ \varrho_\theta^-(-\log \mu(X^n)) \} + \mathbf{E} \{ \Delta_\theta(-\log \mu(X^n)) \} \\
&\leq \frac{1}{2} + \theta + \sum_{|m|=1}^{N_0} [a_m(\theta) + b_m(\theta)] \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \mathbf{E} \{ e^{-2\pi i m \log \mu(X^n)} \} + \epsilon \\
&= \frac{1}{2} + \theta + \sum_{|m|=1}^{N_0} [a_m(\theta) + b_m(\theta)] \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \sum_{j=1}^r \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m} \\
&\quad + \epsilon.
\end{aligned} \tag{37}$$

Let us define now

$$\begin{aligned}
\gamma_n(\epsilon, \theta) &\triangleq \sum_{|m|=1}^{N_0} [|a_m(\theta)| + |b_m(\theta)|] \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \sum_{j: |\lambda_{j,m}| < 1} |\lambda_{j,m}|^{n-1} |\mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m}| \\
&\quad + \epsilon + \theta.
\end{aligned} \tag{38}$$

and recall that N_0 depends on ϵ and θ . Obviously, for every fixed ϵ and θ , the double sum over m and j , in the expression of $\gamma_n(\epsilon, \theta)$, tends to zero as $n \rightarrow \infty$ since all terms contain a factor $|\lambda_{j,m}|^{n-1}$ and by definition of these terms, only $|\lambda_{j,m}| < 1$ are included in the summation. This means that if we let ϵ and θ tend to zero slowly enough with n , thus denoting them by ϵ_n and θ_n , we have $\gamma_n(\epsilon_n, \theta_n) \rightarrow 0$. In particular, let us define ϵ_n and θ_n to be the minimizers³ of $\gamma_n(\epsilon, \theta)$. Then, obviously, $\gamma_n \triangleq \gamma_n(\epsilon_n, \theta_n) \rightarrow 0$ as $n \rightarrow \infty$. Then, our upper and lower bounds become

$$R_n \geq \frac{1}{2} + \sum_{|m|=1}^{N_0} a_m(\theta_n) \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \cdot \sum_{j: |\lambda_{j,m}|=1} \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m} - \gamma_n, \tag{39}$$

and

$$R_n \leq \frac{1}{2} + \sum_{|m|=1}^{N_0} [a_m(\theta_n) + b_m(\theta_n)] \cdot \left(1 - \frac{|m|}{N_0 + 1}\right) \sum_{j: |\lambda_{j,m}|=1} \lambda_{j,m}^{n-1} \cdot \mathbf{l}_{j,m}^T \mathbf{1} \cdot \mathbf{c}_m^T \mathbf{r}_{j,m} + \gamma_n. \tag{40}$$

3.4 Criteria for the Convergent and Oscillatory Modes

Considering the derived lower bound and the upper bound on R_n (eqs. (39) and (40)), it is apparent that the key issue that distinguishes between the convergent mode and the oscillatory mode of R_n , is to determine under what conditions the modulus of the dominant

³Note that with this choice, θ_n and ϵ_n depend only on the parameters of the source μ .

eigenvalue, $\lambda_{1,m}$, namely, the *spectral radius* of A_m , denoted $\rho(A_m)$, is equal to unity and under what conditions it is strictly less than unity (obviously, it cannot be larger than unity). The former case is the oscillatory mode and the latter case is the convergent one. To this end, the following lemma, that appears in [8] (with minor modifications in its phrasing), and that has already been used in earlier related studies [9], [11], proves useful.

Lemma 1 [8, Theorem 8.4.5, p. 509] *Let $F = \{f_{kj}\}$ and $G = \{g_{kj}\}$ be two $r \times r$ matrices. Assume that F is a real, non-negative and irreducible matrix, G is a complex matrix, and $f_{kj} \geq |g_{kj}|$ for all $k, j \in \{1, 2, \dots, r\}$. Then, $\rho(G) \geq \rho(F)$ with equality if and only if there exist real numbers s , and w_1, \dots, w_r such that $G = e^{2\pi i s} D F D^{-1}$, where $D = \text{diag}\{e^{2\pi i w_1}, \dots, e^{2\pi i w_r}\}$.*

The proof of the necessity of the condition $G = e^{2\pi i s} D F D^{-1}$ appears in [8] (see also [9], [11]). The sufficiency is obvious since the matrix $D F D^{-1}$ is similar to F and hence has the same set of eigenvalues.

We wish to apply Lemma 1 in order to distinguish between the two aforementioned cases concerning the spectral radius of A_m . Consider the state transition probability matrix P in the role of F of Lemma 1 (i.e., $f_{kj} = p(j|k)$) and the matrix A_m in the role of G . Since P is assumed positive in this part, then it is obviously non-negative and irreducible. Since it is a stochastic matrix, its spectral radius is, of course, $\rho(P) = 1$. Also, by definition of A_m , as the matrix $\{p(j|k) \cdot \exp[-2\pi i m \log p(j|k)]\}$, it is obvious that the elements of P are the absolute values of the corresponding elements of A_m , and so, all the conditions of Lemma 1 clearly apply. The lemma then tells us that $\rho(A_m) = \rho(P) = 1$ if and only if there exist real numbers s and w_1, \dots, w_r such that:

$$-m \log p(j|k) = (s + w_k - w_j) \bmod 1, \quad j, k = 1, \dots, r, \quad (41)$$

where $x = y \bmod 1$ means that the fractional parts of x and y are equal, that is, $\langle x \rangle = \langle y \rangle$.

To find a vector $\mathbf{w} = (w_1, \dots, w_r)$ and a number s with this property (if exist), we take the following approach: Consider first the choice $k = j$ in (41). This immediately tells us that s , if exists, must be equal to $-m \log p(j|j) \bmod 1$ for every $j = 1, \dots, r$. In other words, one set of conditions is that $-m \log p(j|j)$ are all equal $\bmod 1$, or equivalently,

$$\left\langle m \log \frac{p(j|j)}{p(1|1)} \right\rangle = 0, \quad j = 2, 3, \dots, r, \quad (42)$$

and then s is taken to be the common value of all $\langle -m \log p(j|j) \rangle$. Thus, eq. (41) becomes

$$m \log \frac{p(j|j)}{p(j|k)} = (w_k - w_j) \bmod 1, \quad j, k = 1, \dots, r, \quad (43)$$

and it remains to find the vector \mathbf{w} if possible. To this end, observe that if \mathbf{w} satisfies (43), then for every constant c , $\mathbf{w} + c$ also satisfies (43). Taking $c = -w_1$,⁴ it is apparent that if (43) can hold for some \mathbf{w} , then there is such a vector whose first component vanishes, and then by setting $k = 1$ in (43), we learn that

$$w_j = \left\langle m \log \frac{p(j|1)}{p(j|j)} \right\rangle, \quad j = 1, \dots, r, \quad (44)$$

is a legitimate choice. Thus, (43) becomes

$$\left\langle m \log \left[\frac{p(j|1)p(j|j)}{p(k|1)p(j|k)} \right] \right\rangle = 0 \quad j, k = 1, \dots, r. \quad (45)$$

Note that by setting $k = 1$ in (45), we get (42) as a special case, which means that (45), applied to all $j, k \in \{1, 2, \dots, r\}$, are all the necessary and sufficient conditions needed for $\rho(A_m) = 1$. Now, a necessary and sufficient condition for eq. (45) to hold for *some* integer m , is that the numbers

$$\alpha_{jk} = \log \left[\frac{p(j|1)p(j|j)}{p(k|1)p(j|k)} \right] \quad (46)$$

would be all rational.

We next prove the asymptotic expressions for R_n , first, for the case where some $\{\alpha_{jk}\}$ are irrational, which means that $\rho(A_m) < 1$ for all $m \neq 0$ (convergent mode), and then for the case where all $\{\alpha_{jk}\}$ are rational, which means that there are non-zero values of m for which $\rho(A_m) = 1$ (oscillatory mode).

3.5 Bounds on R_n in the Convergent and Oscillatory Modes

When some α_{jk} are irrational, then for all $m \neq 0$ and $j \in \{1, 2, \dots, r\}$, we have $|\lambda_{j,m}| < 1$, and so, the second terms (i.e., the sums over m) in eqs. (39) and (40) do not exist. Consequently, we immediately get $R_n \geq \frac{1}{2} - \gamma_n$ and $R_n \leq \frac{1}{2} + \gamma_n$, namely, $R_n = \frac{1}{2} + o(1)$.

Consider now the case where all $\{\alpha_{jk}\}$ are rational, and so, there exist $m \neq 0$ with $\rho(A_m) = 1$. Our first step is to establish the fact that if M is the smallest positive integer m that satisfies (45), then any other non-zero integer m satisfies this property if and only if it is an integral multiple of M . The fact that integer multiples of M satisfy (45) is obvious

⁴The choice of the first component of \mathbf{w} is arbitrary.

since $\langle k \cdot M\alpha_{jk} \rangle = \langle k \cdot \langle M\alpha_{jk} \rangle \rangle = \langle k \cdot 0 \rangle = 0$. To see why the converse is true as well, let M' be another integer satisfying (45). If M' is not an integer multiple of M , it must be larger than M since M was defined as the smallest integer satisfying (45). Now, if M and M' both satisfy (45), then so does $M'' = M' - \lfloor M'/M \rfloor \cdot M$, but M'' must be strictly smaller than M , which is a contradiction.

This means that for $m = \ell M$, $\ell = \pm 1, \pm 2, \dots$, and only for these integers, A_m has a modulus 1 eigenvalue

$$\lambda_{1,\ell M} = \exp[2\pi i \langle -\ell M \log p(1|1) \rangle] = \exp[-2\pi i \ell M \log p(1|1)] \quad (47)$$

and the corresponding vector \mathbf{w} is ℓ times (mod 1) the vector \mathbf{w} associated with $m = M$. By the Perron–Frobenius theorem [8], all other eigenvalues have modulus strictly less than 1, and they will contribute exponentially small terms to R_n . Since $\lambda_{1,\ell M}^{-1} A_{\ell M}$ is similar to P , under the transformation matrix $D = \text{diag}\{e^{2\pi i w_1}, \dots, e^{2\pi i w_r}\}$, $w_j = \langle \ell M \log[p(j|1)/p(j|j)] \rangle$, $j = 1, 2, \dots, r$ (see Lemma 1), then by (44), the right- and left eigenvectors associated with $\lambda_{1,\ell M}$ are, respectively,

$$\mathbf{r}_{1,\ell M} = D \cdot \mathbf{1} = \left(1, e^{2\pi i \ell M \log[p(2|1)/p(2|2)]}, \dots, e^{2\pi i \ell M \log[p(r|1)/p(r|r)]}\right)^T, \quad (48)$$

and

$$\mathbf{l}_{1,\ell M} = (\pi_1, \dots, \pi_k) \cdot D^{-1} = \left(\pi_1, \pi_2 e^{-2\pi i \ell M \log[p(2|1)/p(2|2)]}, \dots, \pi_r e^{-2\pi i \ell M \log[p(r|1)/p(r|r)]}\right). \quad (49)$$

Thus, the dominant term in $\mathbf{c}_{\ell M}^T A_{\ell M}^{n-1} \mathbf{1}$ becomes:

$$\lambda_{1,\ell M}^{n-1} \cdot \mathbf{l}_{1,\ell M}^T \mathbf{1} \cdot \mathbf{c}_{\ell M}^T \mathbf{r}_{1,\ell M} = \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)}, \quad (50)$$

where $\zeta_{jk}(n)$ is defined as in Theorem 1. Combining this relation with eq. (39), R_n is further lower bounded as follows:

$$\begin{aligned} R_n &\geq \frac{1}{2} + \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_{\ell M}(\theta_n) \cdot \left(1 - \frac{|\ell M|}{N_0 + 1}\right) \cdot \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)} - \gamma_n \\ &= \frac{1}{2} + \frac{1}{M} \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_{\ell}(M\theta_n) \cdot \left(1 - \frac{|\ell M|}{N_0 + 1}\right) \cdot \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)} - \gamma_n \\ &= \frac{1}{2} + \frac{1}{M} \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_{\ell}(M\theta_n) \cdot \left(1 - \frac{|\ell|}{\lfloor N_0/M \rfloor + 1}\right) \cdot \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)} - \end{aligned}$$

$$\begin{aligned}
& \frac{1}{M} \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_\ell(M\theta_n) \cdot \left[\frac{|\ell|}{(N_0+1)/M} - \frac{|\ell|}{\lfloor N_0/M \rfloor + 1} \right] \cdot \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)} - \gamma_n \\
\geq & \frac{1}{2} + \frac{1}{M} \sum_{j,k} p_j \pi_k \left\{ \varrho_{M\theta_n}^{\bar{}}[\zeta_{jk}(n)] - \frac{1}{2} - \eta_n \right\} - \\
& \frac{1}{M} \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_\ell(M\theta_n) \cdot \left[\frac{|\ell|}{(N_0+1)/M} - \frac{|\ell|}{\lfloor N_0/M \rfloor + 1} \right] \cdot \sum_{j,k} p_j \pi_k e^{2\pi i \ell \zeta_{jk}(n)} - \gamma_n \\
\geq & \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \sum_{j,k} p_j \pi_k \varrho[\zeta_{jk}(n)] - \frac{1}{M} \sum_{j,k} p_j \pi_k \Delta_{M\theta_n}[\zeta_{jk}(n)] - \\
& \frac{1}{M} \sum_{j,k} p_j \pi_k \sum_{|\ell|=1}^{\lfloor N_0/M \rfloor} a_\ell(M\theta_n) \cdot \left[\frac{|\ell|}{(N_0+1)/M} - \frac{|\ell|}{\lfloor N_0/M \rfloor + 1} \right] \cdot e^{2\pi i \ell \zeta_{jk}(n)} - \\
& \gamma_n - \frac{\eta_n}{M}, \tag{51}
\end{aligned}$$

where η_n is defined as the maximum approximation error of the function $\varrho_{M\theta_n}^{\bar{}}$ using $\lfloor N_0(\epsilon_n, \theta_n)/M \rfloor$ terms of the Fejér series. We wish to show now that $\eta_n \rightarrow 0$ as $n \rightarrow \infty$. Let us assume that ϵ_n and θ_n are small enough to make $N_0 = N_0(\epsilon_n, \theta_n)$ not smaller than $2M$, and so, $\lfloor N_0/M \rfloor \geq N_0/M - 1 \geq N_0/2M$. Then, using eq. (26),

$$\begin{aligned}
\eta_n & \leq \epsilon_0 \left[\frac{N_0(\epsilon_n, \theta_n)}{2M}, M\theta_n \right] \\
& = \inf_{0 < \delta < 1/2} \left[\frac{\delta}{M\theta_n} + \frac{2M}{N_0(\epsilon_n, \theta_n) \sin^2(\pi\delta)} \right] \\
& < \inf_{0 < \delta < 1/2} \left[\frac{2M\delta}{\theta_n} + \frac{2M}{N_0(\epsilon_n, \theta_n) \sin^2(\pi\delta)} \right] \\
& = 2M \cdot \epsilon_0[N_0(\epsilon_n, \theta_n), \theta_n] \\
& \leq 2M\epsilon_n \rightarrow 0, \tag{52}
\end{aligned}$$

where the last inequality follows from eq. (27). Thus, η_n/M in the last line of (51), is upper bounded by $2\epsilon_n$. The first two terms in the last expression of (51) form Ω_n , as defined in Theorem 1. Now, for the absolute value of the fourth term, it is first observed that upon a standard algebraic manipulation under the assumption $N_0 \geq 2M$, we have

$$\begin{aligned}
\left| \frac{1}{(N_0+1)/M} - \frac{1}{\lfloor N_0/M \rfloor + 1} \right| & = \frac{M | \langle N_0/M \rangle + 1/M - 1 |}{(N_0+1)(\lfloor N_0/M \rfloor + 1)} \\
& \leq \frac{2M^2}{N_0^2}. \tag{53}
\end{aligned}$$

Thus, the fourth term of (51) is upper bounded by the weighted sum (with weights $p_j \pi_k$

for each pair (j, k) of terms, that are bounded as follows:

$$\begin{aligned}
& \left| \frac{1}{M} \sum_{|\ell|=1}^{N_0/M} a_\ell(M\theta_n) \cdot \left[\frac{|\ell|}{(N_0+1)/M} - \frac{|\ell|}{N_0/M+1} \right] \cdot e^{2\pi i \ell \zeta_{jk}(n)} \right| \\
& \leq \frac{4M^2}{N_0^2} \sum_{\ell=1}^{\lfloor N_0/M \rfloor} \ell \cdot |a_\ell(M\theta_n)| \\
& = \frac{2M}{\pi N_0^2} \sum_{\ell=1}^{\lfloor N_0/M \rfloor} \frac{|1 - e^{-2\pi i \ell M\theta_n}|}{2\pi \ell M\theta_n} \\
& = \frac{2M}{\pi N_0^2} \sum_{\ell=1}^{\lfloor N_0/M \rfloor} \sqrt{\frac{2[1 - \cos(2\pi \ell M\theta_n)]}{(2\pi \ell M\theta_n)^2}}. \tag{54}
\end{aligned}$$

To bound the summand of the last expression, consider the following: For every positive t , clearly, $\sin t \leq t$, and so, for every $\alpha > 0$,

$$1 - \cos \alpha = \int_0^\alpha \sin t dt \leq \int_0^\alpha t dt = \frac{\alpha^2}{2}, \tag{55}$$

which for $\alpha = 2\pi \ell M\theta_n$, implies that the summand is bounded by 1, and hence the expression in the last chain of inequalities is further upper bounded by $\delta_n \triangleq 2/(\pi N_0)$. Since $N_0 = N_0(\epsilon_n, \theta_n) \rightarrow \infty$, then $\delta_n \rightarrow 0$, and we have

$$\begin{aligned}
R_n & \geq \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \sum_{j,k} p_j \pi_k \varrho[\zeta_{jk}(n)] - \frac{1}{M} \sum_{j,k} p_j \pi_k \Delta_{M\theta_n}[\zeta_{jk}(n)] - \gamma_n - 2\epsilon_n - \delta_n \\
& \geq \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \sum_{j,k} p_j \pi_k \varrho[\zeta_{jk}(n)] - \\
& \quad \frac{1}{M} \sum_{j,k} p_j \pi_k \mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (M\theta_n, 1 - M\theta_n)\} - \gamma_n - 2\epsilon_n - \delta_n, \tag{56}
\end{aligned}$$

and so, the lower bound of Theorem 1 is obtained with $\xi_n \triangleq M\theta_n$. In the very same manner, the upper bound on R_n is given by

$$\begin{aligned}
R_n & \leq \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \sum_{j,k} p_j \pi_k \varrho_{\bar{M}\theta_n}[\zeta_{jk}(n)] + \\
& \quad \frac{1}{M} \sum_{j,k} p_j \pi_k \Delta_{M\theta_n}[\zeta_{jk}(n)] + \gamma_n + 2\epsilon_n + \delta_n \tag{57}
\end{aligned}$$

$$\begin{aligned}
& \leq \frac{1}{2} \left(1 - \frac{1}{M} \right) + \frac{1}{M} \sum_{j,k} p_j \pi_k \varrho[\zeta_{jk}(n)] + \\
& \quad \frac{1}{M} \sum_{j,k} p_j \pi_k \mathcal{I}\{\varrho[\zeta_{jk}(n)] \notin (\xi_n, 1 - \xi_n)\} + \gamma_n + 2\epsilon_n + \delta_n, \tag{58}
\end{aligned}$$

which is the upper bound of Theorem 1. Here, one has to bound also an expression similar to (54), but with $a_\ell(M\theta_n)$ being replaced by $b_\ell(M\theta_n)$, and the bounding technique is similar. This completes the proof of Theorem 1.

4 Extensions

We now discuss some extensions of Theorem 1. In particular, we drop the assumption that all transition probabilities must be strictly positive and first assume that P corresponds to an irreducible aperiodic Markov source. Then we drop the aperiodicity constraint.

4.1 Irreducible Aperiodic Markov Sources

When some of the entries of the matrix P vanish, then obviously, Theorem 1 cannot be used as is since the corresponding parameters α_{jk} are no longer well defined. Lemma 1, which stands at the heart of the proof of Theorem 1, can still be used as long as P is irreducible, but more caution should be exercised. The key issue is still to determine whether there exist parameters s and \mathbf{w} (and to find them if exist) that satisfy

$$-m \log p(j|k) = (s + w_k - w_j) \bmod 1, \quad (59)$$

but now these equations are imposed only for the pairs (j, k) for which $p(j|k) > 0$ (as for the other pairs $a_{jk}(m) = p(j|k) = 0$ satisfy the conditions of Lemma 1 automatically anyway). The approach taken in the solution for s and \mathbf{w} , that was derived in the first part of Section 3, can still be applied, with some minor modifications, as long as at least some particular subsets of the entries of P are still positive.

For example, if one or more diagonal element of P is positive, and for all positive $p(j|j)$, the numbers $\langle -m \log p(j|j) \rangle$ are equal, then s can still be taken to be the common value of all these numbers. If, in addition, at least one row of P is strictly positive, say, row number l , then w_j can be taken to be $\langle m \log [p(l|l)/p(j|l)] \rangle$, and then the rationality condition of Theorem 1 is replaced by the condition that

$$\alpha'_{jk} = \log \left[\frac{p(j|l)p(l|l)}{p(k|l)p(j|k)} \right] \quad (60)$$

must be rational for all (j, k) with $p(j|k) > 0$. The bounds on R_n in the oscillatory mode would be exactly as in Theorem 1, but with the above assignments of s and \mathbf{w} .

For a general non-negative matrix P , however, it may not be a trivial task to determine whether equations (59) have a solution, and if so, what this solution is. In fact, it may be simpler and more explicit to check directly if A_m has an eigenvalue on the unit circle (which thereby dictates s) and then to find \mathbf{w} using Lemma 1. This would lead to the following generalized version of Theorem 1.

Theorem 2 Consider the Shannon code of block length n for an irreducible aperiodic Markov source. Let M be defined as the smallest positive integer m such that

$$\rho(A_m) \equiv |\lambda_{1,m}| = 1 \quad (61)$$

and as $M = \infty$ if (61) does not hold for any positive integer m . Then, R_n is characterized as follows:

(a) If $M = \infty$, then

$$R_n = \frac{1}{2} + o(1). \quad (62)$$

(b) If $M < \infty$, then the bounds of Theorem 1, part (b), hold with $\zeta_{jk}(n)$ being redefined according to

$$\zeta_{jk}(n) = M[(n-1)s + w_j - w_k - \log p_j], \quad (63)$$

where

$$s = \frac{\arg\{\lambda_{1,M}\}}{2\pi} \quad (64)$$

and

$$w_j = \frac{\arg\{x_j\}}{2\pi}, \quad j = 1, 2, \dots, r, \quad (65)$$

x_j being the j -th component of the right eigenvector \mathbf{x} of A_M , which is associated with the dominant eigenvalue $\lambda_{1,M}$.

The proof of Theorem 2 is very similar to that of Theorem 1, and hence we will not provide it here. In a nutshell, we observe that the Perron–Frobenius Theorem and Lemma 1 are still applicable. Then, we use the necessity of the condition $A_m = e^{2\pi is} D P D^{-1}$ and the fact that once this condition holds, the vector $\mathbf{x} = D \cdot \mathbf{1} = (e^{2\pi i w_1}, \dots, e^{2\pi i w_r})^T$ is the right eigenvector associated with the dominant eigenvalue $\lambda_{1,m} = e^{2\pi is}$.

Unfortunately, Theorem 2 does not suggest a practical way to find M . One must start with $m = 1$, check if $\rho(A_1) = 1$; if not – increment m to 2, check $\rho(A_2)$, and so on. In the event that $M = \infty$, we do not have a stopping rule and we may keep incrementing m indefinitely. An interesting point to note, however, is that the oscillatory expression goes to $1/2$ when M grows without bound. This means that given the block length n , it is sufficient to stop incrementing m at some $m(n)$, where $m(n)$ is an arbitrary function that grows (and no matter how slowly) with n . This is because the oscillatory expression will then be $1/2 + o(n)$ anyway, just like the convergent expression, so the distinction between the two modes loses its meaning.

Finally, it is instructive to demonstrate an example of a *reducible* Markov source, for which Theorems 1 and 2 do not hold, and see that even in a simple situation ($r = 2$), once the irreducibility assumption is dropped, the two-mode behavior, predicted by Theorems 1 and 2, disappears. Thus, the point in Example 2 below is that the irreducibility assumption is not imposed just for technical convenience. It is actually essential for Theorems 1 and 2 to hold.

Example 2. *Reducible Markov source.* Consider the case $r = 2$, where $p(1|2) = 0$ and $\alpha \triangleq p(2|1) \in (0, 1)$, i.e.,

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ 0 & 1 \end{pmatrix}. \quad (66)$$

Assume also that $p_1 = 1$ and $p_2 = 0$. Since this is a reducible Markov source (once in state 2, there is no way back to state 1), we cannot use Theorems 1 and 2, but we can still find an asymptotic expression of the redundancy in a direct manner: Note that the chain starts at state ‘1’ and remains there for a random duration, which is a geometrically distributed random variable with parameter $(1 - \alpha)$. Thus, the probability of k 1’s (followed by $n - k$ 2’s) is about $(1 - \alpha)^k \cdot \alpha$ (for large n) and so the argument of the function $\varrho(\cdot)$ should be the negative logarithm of this probability. Taking the expectation w.r.t. the randomness of k , we readily have

$$R_n = \sum_{k=0}^{\infty} \alpha(1 - \alpha)^k \varrho[-\log \alpha - k \log(1 - \alpha)] + o(1). \quad (67)$$

We see then that there is *no oscillatory mode* in this case, as R_n always tends to a constant that depends on α , in contrast to the convergent mode of Theorems 1 and 2, where the limit is always $1/2$, independently of the source statistics. To summarize, it is observed that the behavior here is very different from that of the irreducible case, characterized by Theorems 1 and 2.

4.2 Irreducible Periodic Markov Sources

Consider now an irreducible periodic Markov source. The Perron-Frobenius theorem and Lemma 1 still hold [8]. However, the matrix P now has d eigenvalues on the unit circle, namely, all the d -th roots of unity [8], where d is the period, i.e.,

$$\lambda'_t = e^{2\pi i t/d}, \quad t = 0, 1, \dots, d - 1. \quad (68)$$

Let \mathbf{r}_t and \mathbf{l}_t be the right- and the left eigenvectors of P that are associated with λ'_t . The analysis is similar as in the aperiodic case, except that we now have d oscillatory terms, one for each eigenvalue on the unit circle. Indeed, suppose that for some m , the matrix A_m has a modulus-1 eigenvalue $\lambda = e^{2\pi i s}$. Then, of course,

$$A'_m \triangleq e^{-2\pi i s} A_m \quad (69)$$

has eigenvalue 1. By definition, the entries of P are still the absolute values of the corresponding entries of A'_m , as in Lemma 1. Thus, by this lemma, A'_m is similar to P , and so it has the same eigenvalues as P . Among them, the d -th roots of unity λ'_t , $t = 0, 1, \dots, d-1$ are eigenvalues of A'_m . Therefore, A_m has the following eigenvalues on the unit circle:

$$\lambda_{t,m} = e^{2\pi i(s+t/d)}, \quad t = 0, 1, \dots, d-1. \quad (70)$$

Let us relabel, if necessary, the eigenvalues of A_m such that $s \in [0, 1/d)$. This means that the definition of s in Theorem 2 should be restricted to the half open interval $[0, 1/d)$. Thus, Theorem 2 holds except that $\zeta_{jk}(n)$ are replaced by

$$\zeta_{jkt}(n) \triangleq M \left[(n-1) \left(s + \frac{t}{d} \right) + w_j - w_k - \log p_j \right], \quad j, k \in \{1, 2, \dots, r\}, \quad t = 0, 1, \dots, d-1 \quad (71)$$

and the double summations over (j, k) with weights $p_j \pi_k$, are replaced by corresponding triple summations over (j, k, t) with weights $p_j r_{t,j} l_{t,k}$, where $l_{t,k}$ is the k -th component of \mathbf{l}_t and $r_{t,j}$ is the j -th component of \mathbf{r}_t . Note that $r_{0,j} = 1$ and $l_{0,k} = \pi_k$, so for $d = 1$ we indeed obtain the expression (63) of the aperiodic case as a special case.

Appendix

In this appendix, we establish the relation (26). As is shown in [20], the coefficients of the N -th order Fejér series expansion, $\{f(u)\}_N$, of a general periodic function $f(u)$, with period 1, are given by the Fourier coefficients f_m multiplied by the “triangular window” $1 - |m|/(N+1)$. This means that in the original u -domain, the reconstruction $\{f(u)\}_N$ is given by the convolution between $f(u)$ and the kernel

$$K_N(u) = \sum_{m=-N}^N \left(1 - \frac{|m|}{N+1} \right) e^{2\pi i m u} = \frac{\sin^2[(N+1)\pi u]}{(N+1) \sin^2(\pi u)}. \quad (\text{A.1})$$

Since $\int_{-1/2}^{+1/2} K_N(u) du = 1$, we have

$$\begin{aligned}
|f(u) - \{f(u)\}_N| &= \left| f(u) - \int_{-1/2}^{+1/2} dt f(u-t) K_N(t) \right| \\
&= \left| \int_{-1/2}^{+1/2} dt [f(u) - f(u-t)] K_N(t) \right| \\
&\leq \int_{-1/2}^{+1/2} dt |f(u) - f(u-t)| \cdot K_N(t) \\
&= \int_{|t| \leq \delta} dt |f(u) - f(u-t)| \cdot K_N(t) + \\
&\quad \int_{\delta \leq |t| \leq 1/2} dt |f(u) - f(u-t)| \cdot K_N(t) \tag{A.2}
\end{aligned}$$

for every $\delta \in (0, 1/2)$. Now, in our case, for all three functions, $|t| \leq \delta$ implies $|f(u) - f(u-t)| \leq \delta/\theta$, since the maximum absolute slope of all three of them is $1/\theta$. Since $K_N(t) \geq 0$ and $\int_{-1/2}^{1/2} dt K_N(t) = 1$, the first integral in the last line is bounded by δ/θ . As for the second integral, in our case, $|f(u) - f(u-t)| \leq 1$ for all three functions. Since the sine function is monotonically increasing in the range $[0, \pi/2]$, then $1/2 \geq |t| \geq \delta$ implies

$$K_N(t) \leq \frac{1}{(N+1) \sin^2(\pi\delta)} < \frac{1}{N \sin^2(\pi\delta)}. \tag{A.3}$$

Thus, for every $\delta \in (0, 1/2)$,

$$|f(u) - \{f(u)\}_N| \leq \frac{\delta}{\theta} + \frac{1}{N \sin^2(\pi\delta)} \tag{A.4}$$

and eq. (26) is obtained upon minimizing the r.h.s. over the free parameter δ .

References

- [1] J. Abrahams, “Code and parse trees for lossless source encoding,” *Communications in Information and Systems*, vol. 1, no. 2, pp. 113–146, April 2001.
- [2] D. Blackwell and J. L. Hodges, “The probability in the extreme tail of a convolution,” *Ann. Math. Stat.*, vol. 30, pp. 1113–1120, 1959.
- [3] M. Drmota and W. Szpankowski, “Precise minimax redundancy and regret,” *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2686–2707, November 2004.
- [4] C.-G. Esseen, “Fourier analysis of distribution functions,” *Acta Mathematica*, vol. 77, pp. 1-125, 1945.

- [5] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.
- [6] R. G. Gallager, “Variations on the theme by Huffman,” *IEEE Trans. Inform. Theory*, vol. IT-24, no. 6, pp. 668–674, November 1978.
- [7] R. M. Gray, “Quantization noise spectra,” *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1220–1244, November 1990.
- [8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [9] P. Jacquet, W. Szpankowski, and J. Tang, “Average profile of the Lempel–Ziv parsing scheme for a Markovian source,” *Algorithmica*, vol. 31, pp. 318–360, 2001.
- [10] P. Jacquet and W. Szpankowski, “Markov types and minimax redundancy for Markov sources,” *IEEE Trans. Inform. Theory*, vol. 50, no. 7, pp. 1393–1402, July 2004.
- [11] P. Jacquet and W. Szpankowski, “Joint string complexity for Markov sources,” *23rd International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, AofA’12, DMTCS Proc.*, pp. 303–322, Montreal, 2012.
- [12] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley, New York, 1974.
- [13] G. Louchard and W. Szpankowski, “Average redundancy of the Lempel–Ziv code,” *IEEE Trans. Inform. Theory*, vol. 43, no. 1, pp. 2–8, January 1997.
- [14] N. Merhav, “Relations between redundancy patterns of the Shannon code and wave diffraction patterns of partially disordered media,” *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3402–3406, June 2012.
- [15] J. Rissanen, “Complexity of strings in the class of Markov sources,” *IEEE Trans. Inform. Theory*, vol. IT-32, no. 4, pp. 526–532, July 1986.
- [16] S. A. Savari, “Variable-to-fixed length codes for predictable sources,” *Proc. Data Compression Conference (DCC)*, Snowbird, UT, pp. 481–490, 1998.

- [17] S. A. Savari and R. G. Gallager, “Generalized Tunstall codes for sources with memory,” *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 658–668, March 1997.
- [18] W. Szpankowski, “Asymptotic average redundancy of Huffman (and other) block Codes,” *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2434–2443, November 2000.
- [19] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, John Wiley & Sons, New York, 2001.
- [20] B. Sury, “Weierstrass’s Theorem – Leaving No Stone Unturned,” *Workshop on Linear Algebra and Analysis*, University of Hyderabad, 2006.
<http://www.isibang.ac.in/~sury/hyderstone.pdf>