Average Redundancy of the Shannon Code for Markov Sources

Neri Merhav and Wojciech Szpankowski

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

and

Department of Computer Science

Purdue University,

W. Lafayette, IN U.S.A.

ISIT 2013, Istanbul, Turkey, July 2013.

Background

Redundancy rates of lossless codes have received the attention of many researchers:

A very partial list includes:

Krichevsky (1968), Gallager (1978), Rissanen (1986), Capocelli and De Santis (1992), Savari and Gallager (1997), Savari (1998), Szpankowski (2000), Jacquet and Szpankowski (2001, 2012), Abrahams (2001), Drmota and Szpankowski (2004), Merhav (2012).

Background

Redundancy rates of lossless codes have received the attention of many researchers:

A very partial list includes:

Krichevsky (1968), Gallager (1978), Rissanen (1986), Capocelli and De Santis (1992), Savari and Gallager (1997), Savari (1998), Szpankowski (2000), Jacquet and Szpankowski (2001, 2012), Abrahams (2001), Drmota and Szpankowski (2004), Merhav (2012).

Background

Redundancy rates of lossless codes have received the attention of many researchers:

A very partial list includes:

Krichevsky (1968), Gallager (1978), Rissanen (1986), Capocelli and De Santis (1992), Savari and Gallager (1997), Savari (1998), Szpankowski (2000), Jacquet and Szpankowski (2001, 2012), Abrahams (2001), Drmota and Szpankowski (2004), Merhav (2012).

This work is a further development on Szpankowski (2000) and Merhav (2012).

Background (Cont'd)

Theorem 1 (Szpankowski 2000) *Consider the Huffman block code of length n over a binary memoryless source with $p < \frac{1}{2}$ and define*

$$\alpha = \log_2 \left(\frac{1-p}{p} \right)$$

and

$$\beta = \log_2 \left(\frac{1}{p} \right).$$

Then for large n,

$$R_n = \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} + o(1) \approx 0.057304 & \alpha \quad \text{irrational} \\ \\ \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n \rangle - \frac{1}{2} \right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle \beta M n \rangle / M} + o(1) & \alpha = \frac{N}{M} \end{cases}$$

where N and M are relatively prime integers, and $\langle x \rangle = x - \lfloor x \rfloor$.

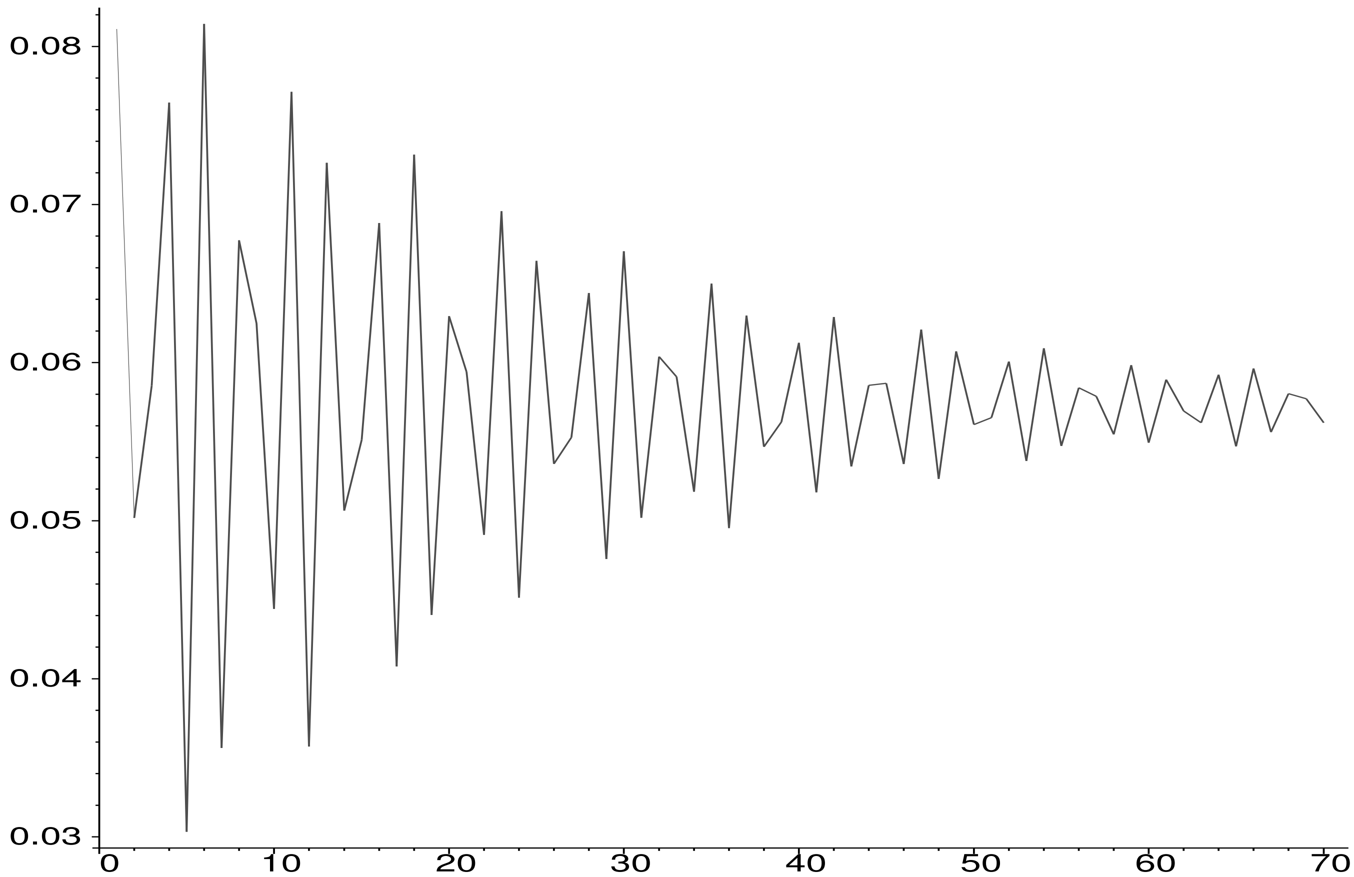

Figure 1: Irrational case $p = 1/\pi$ (Szpankowski 2000).

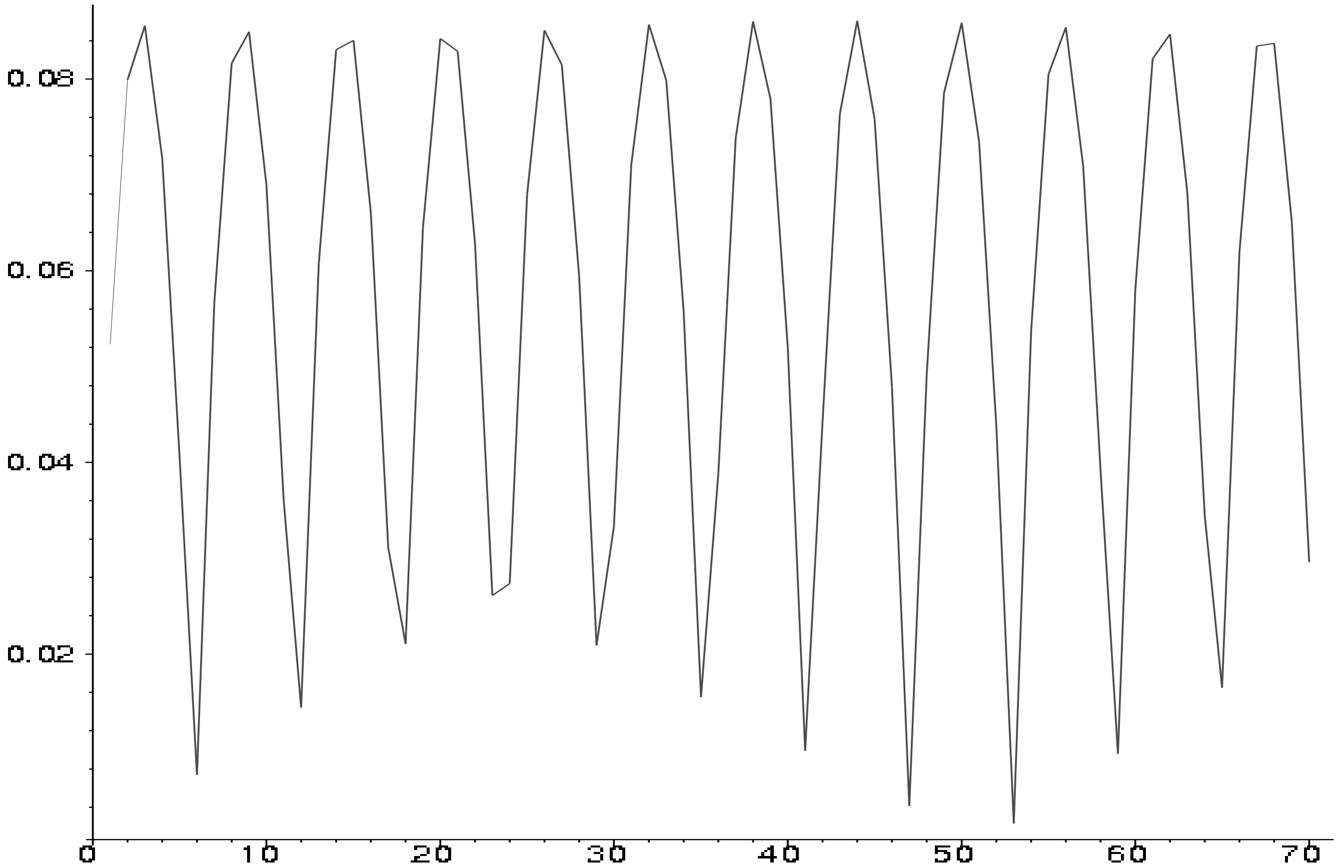

Figure 2: Rational case $p = 1/9$ (Szpankowski 2000).

Comments

- Similar behavior for the Shannon code ($R_n \to 1/2$ in the irrational case.)
- Easy extension to memoryless sources of a general finite alphabet:
  - Rationality condition: $\alpha_j = \log_2 p_j/p_1$ should all be rational.
  - Fundamental frequency of oscillations: $\beta = -\log p_1$.
- Merhav (2012): relation to wave diffraction in disordered media (Bragg peaks).

This Work – Extension to the Markov Case

- Here too: two–mode behavior, depending on some rationality conditions.
- Non–trivial extension in the following respects:
  - Analysis tools.
  - Rationality conditions.
  - Oscillatory mode redundancy expressions.
  - Strong dependence of the “dynamics”: ir/reducibility, a/periodicity.
- Applicability of analysis method to other codes (Huffman included).
- Applicability to other problem areas:
  - Uniform quantization.
  - Bragg peaks in disordered media.
  - Statistics of round–off errors of decoding metrics with finite–precision arithmetics.

Markov Sources

Source sequence $X_1, X_2, \ldots$ over alphabet $\mathcal{X} = \{1, 2, \ldots, r\}$ is generated by a first–order Markov chain with a given matrix

$$P = \{p(j|k)\}_{j,k=1}^{r}.$$

with initial state probabilities p_k, $k = 1, 2, \ldots, r$;
stationary state probabilities π_k, $k = 1, 2, \ldots, r$.

For $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ under the given Markov source, is

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

The average redundancy is

$$R_n = \mathbf{E}\{\lceil -\log P(X^n) \rceil + \log P(X^n)\} = \mathbf{E}\{\varrho[-\log P(X^n)]\}.$$

where $\varrho(u) = \lceil u \rceil - u$.

Main Result for Positive Transition Matrices

Theorem 2 *Consider the Shannon code of length n for a Markov source with a positive matrix P. Define*

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

(a) If not all $\{\alpha_{jk}\}$ are rational, then $R_n = \frac{1}{2} + o(1)$.
(b) If all $\{\alpha_{jk}\}$ are rational, then let

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

Then,

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

where M is the smallest common multiple of the denominators of $\{\alpha_{jk}\}$.

Basic Idea of the Proof

Consider the Fourier (Fejer) series expansion of the periodic function $\varrho(u) = \lceil u \rceil - u$:

$$\varrho(u) = \frac{1}{2} + \sum_{m \neq 0} a_m \exp\{2\pi i m u\},$$

so that

$$R_n = \mathbf{E}\varrho[-\log P(X^n)] = \frac{1}{2} + \sum_{m \neq 0} a_m \mathbf{E} \exp\{-2\pi i m \log P(X^m)\}.$$

By the Markov property, $\mathbf{E} \exp\{-2\pi i m \log P(X^n)\}$ is given in terms of A_m^n where A_m is a matrix defined as

$$A_m = \{p(k|j) \exp\{-2\pi i m \log p(k|j)\}_{j,k=1}^{r}.$$

Basic Idea of the Proof (Cont'd)

In particular,

$$\mathbf{E}\exp\{-2\pi i m \log P(X^n)\} = \sum_{\ell=1}^{r} \text{coeff}_\ell \cdot \lambda_{m,\ell}^n.$$

The eigenvalue $\lambda_{m,\ell}$ of A_m with the largest modulus (spectral radius) dominates the m–th term for n large.

- Spectral radius $< 1 \to$ all terms decay and $R_n \to 1/2$.
- Spectral radius $= 1$ for some $m = m_0 \to$ also $\forall$ multiples of m_0, forming the oscillatory terms of R_n.
- Nec. and suff. conditions for spectral radius $= 1$: matrix theory.

Irreducible Aperiodic Case

Theorem 3 *Let* m_0 *be the smallest positive* m *such that* $\rho(A_m) = |\lambda_{1,m}| = 1$*.*
(a) If $m_0 = \infty$*,* $R_n = \frac{1}{2} + o(1)$*.*
(b) If $m_0 < \infty$*, let*

$$s = \frac{\textit{arg}\{\lambda_{1,m_0}\}}{2\pi}, \qquad w_j = \frac{\textit{arg}\{x_j\}}{2\pi},$$

x_j *being the* j*–th coordinate of the right–eigenvector of* A_{m_0} *pertaining to* λ_{1,m_0}*. Then,*

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

with

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

Irreducible Periodic Case

- Let d be the period of the chain.
- Theorem – essentially the same except that P has d dominant eigenvalues – the d–th order roots of 1.
- Each such eigenvalue contributes a double–sum of oscillatory terms with a different fundamental frequency:

$$\zeta_{jkt}(n) = M[(n-1)(s + t/d) + w_j - w_k - \log p_j].$$

Reducible Case

We don't have a general theory here, but here is a simple example showing that the behavior may be entirely different:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{pmatrix}.$$

Assume $p_1 = 1$ and $p_2 = 0$. Direct computation shows that

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

No oscillatory mode, as R_n always tends to a constant that depends on the source.

Thank You!