Asymptotic Average Redundancy of Huffman (and Shannon-Fano) Block Codes*

W. Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

June 13, 2000

*Research supported by NSF Grants NCR-9415491 and C-CR 9804760

Outline

1. The Redundancy Rate Problem: Basic definitions
2. Shannon-Fano Code Redundancy
3. Huffman Code Redundancy
4. Sketch of Proof by Analytic Methods
5. BONUS: Redundancy of Golomb's Code

Definitions

The **redundancy-rate problem** of universal coding for a class of sources consists in determining by how much the actual code length exceeds the optimal (ideal) code length.

A code

$$C_n : \mathcal{A}^n \to \{0,1\}^*$$

is a mapping from the set $\mathcal{A}^n$ of all sequences of length n over the alphabet $\mathcal{A}$ to the set $\{0,1\}^*$ of binary sequences.

Given a probabilistic source model and a code C_n we let:

- $P(x_1^n)$ be the probability of the message $x_1^n = x_1 \ldots x_n$,
- $L(C_n, x_1^n)$ be the code length for x_1^n,
- Entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n) \lg P(x_1^n)$,
- the "ideal" code length: $-\lg P(x_1^n)$.

Information-theoretic quantities are expressed in binary logarithms written $\lg := \log_2$. We also write $\log := \ln$.

Various Redundancies

The **pointwise redundancy** $R_n(C_n, P; x_1^n)$ and the **average redundancy** $\bar{R}_n(C_n, P)$ are defined as

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

where $\mathbf{E}$ denotes the expectation. The **maximal** redundancy is defined as

$$R^*(C_n, P) = \max_{x_1^n}\{R_n(C_n, P; x_1^n)\}.$$

The pointwise redundancy can be negative, maximal and average redundancy cannot (see next slide).

The **strong redundancy-rate problem** consists in determining for a class $\mathcal{S}$ of source models the rate of growth of the minimax quantities

$$\begin{aligned} \bar{R}_n^*(\mathcal{S}) &= \min_{C_n} \max_{P \in \mathcal{S}}\{\bar{R}_n(C_n, P)\}(= o(n)), \\ R_n^*(\mathcal{S}) &= \min_{C_n} \max_{P \in \mathcal{S}}\{R_n^*(C_n, P)\}(= o(n)) \end{aligned}$$

as $n \to \infty$.

Shannon's Lower Bound

Fact: *For any code, the average code length* $\mathbf{E}[L(C_n, X_1^n)]$ *cannot be smaller than the entropy of the source* $H_n(P)$*, that is,*

$$\mathbf{E}[L(C_n, X_1^n)] \geq H_n(P).$$

Sketch of Proof: Let $K = \sum_{x_1^n} 2^{-L(x_1^n)} \leq 1$, and $L(C_n, x_1^n) := L(C_n)$. Then

$$\begin{aligned}
\mathbf{E}[L(C_n, X_1^n)] - H_n(P) &= \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) L(x_1^n) \\
&+ \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) \log P(x_1^n) \\
&= \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) \log \frac{P(x_1^n)}{2^{-L(x_1^n)}/K} - \log K \\
&\geq 0
\end{aligned}$$

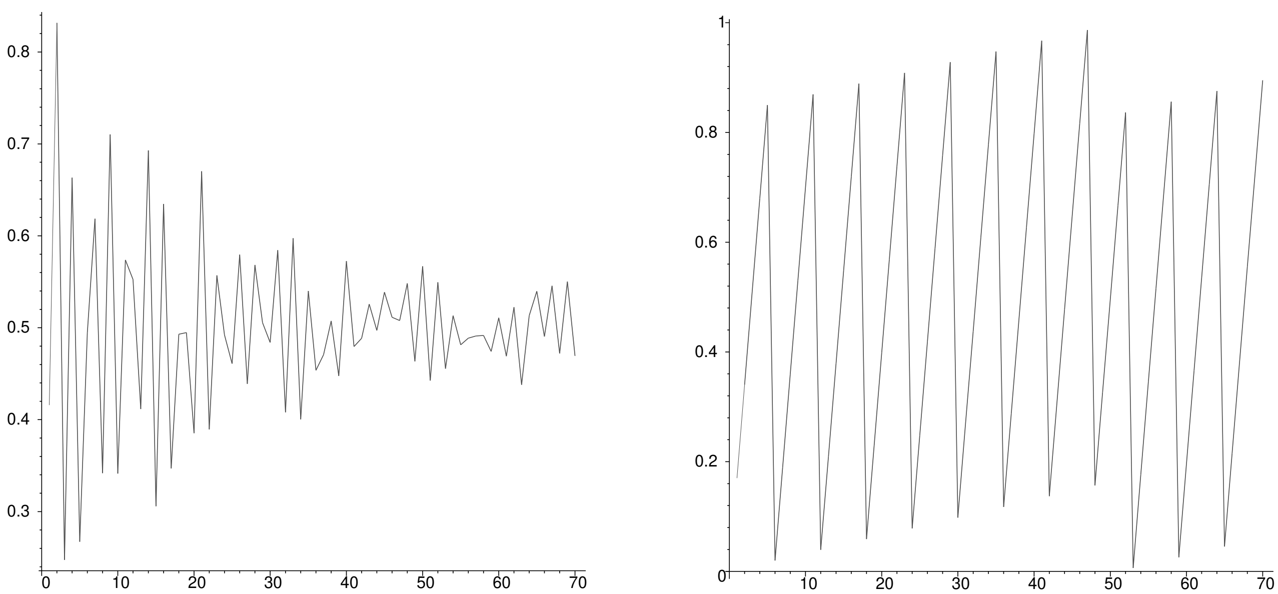
since the first term is a divergence and cannot be negative (or $\log x \leq x - 1$) while $K \leq 1$ by Kraft's inequality.

Shannon-Fano Code

Assume we know the distribution $P(x_1^n)$. Then Shannon (Shannon-Fano) suggested a code that assigns code length $\lceil -\lg P(x_1^n) \rceil$ to x_1^n. Our goal is to compute the average redundancy of such a code.

Consider a binary memoryless source with p denoting the probability of generating 0. For a block of length n, the average redundancy $\bar{R}_n^{SF}$ of the Shannon-Fano code can be computed as follows

$$
\begin{aligned}
\bar{R}_n^{SF} &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \left(\lceil -\log_2 (p^k (1-p)^{n-k}) \rceil \right. \\
&+ \left. \log_2 (p^k (1-p)^{n-k}) \right) .
\end{aligned}
$$



(a) (b)

Figure 1: Shannon–Fano code redundancy versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

Redundancy of Shannon-Fano Code

Let

$$\alpha = \log_2\left(\frac{1-p}{p}\right), \quad \beta = \log_2\left(\frac{1}{1-p}\right).$$

We also denote by

$$\langle x \rangle = x - \lfloor x \rfloor$$

the fractional part of x

Theorem 1. *Consider the Shannon–Fano block code of length n binomially(n,p) distributed over a binary alphabet. Then, for $p < \frac{1}{2}$ as $n \to \infty$*

$$\bar{R}_n^{SF} = \begin{cases} \frac{1}{2} + o(1) & \alpha \quad \text{irrational} \\ \\ \frac{1}{2} - \frac{1}{M}\left(\langle Mn\beta \rangle - \frac{1}{2}\right) + O(\rho^n) & \alpha = \frac{N}{M}, \quad \gcd(N, M) = 1 \end{cases}$$

where $\rho < 1$ and N, M are integers such that $\gcd(N, M) = 1$.

Huffman Code

As before we consider a binary memoryless source emitting 0 and 1 with probabilities p and $q = 1 - p$, respectively. For a block of length n, we construct its **Huffman code** (through the associated Huffman tree).

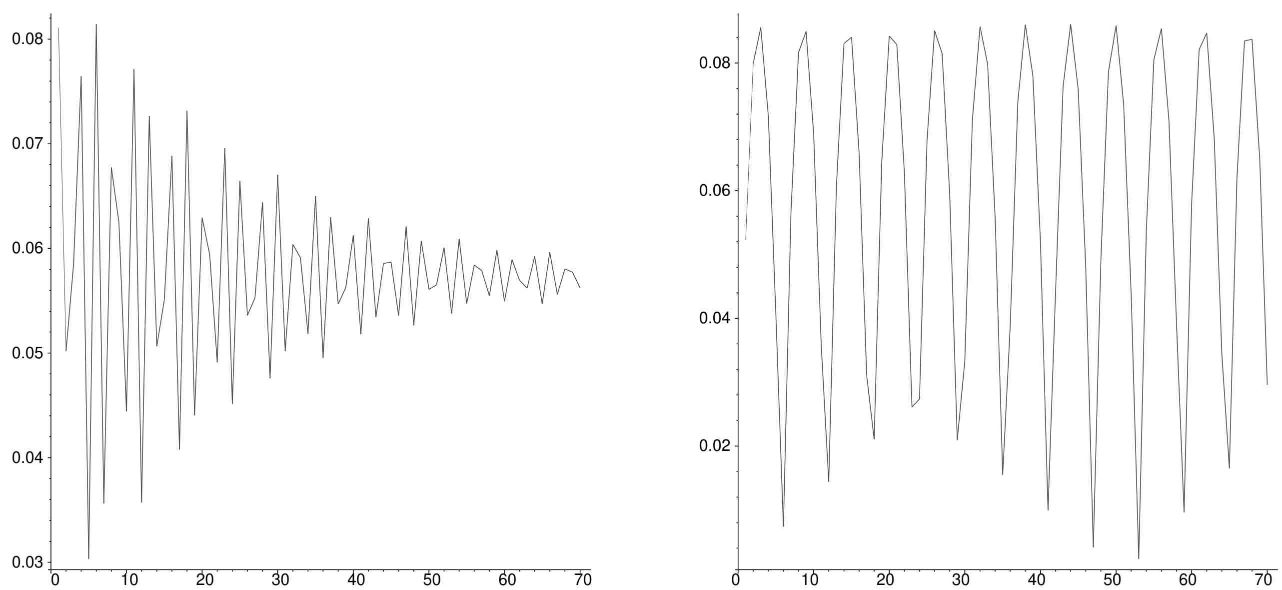
Using Stubley's result (1994), we conclude that the average redundancy $\bar{R}_n^H$ of the Huffman code is

$$\bar{R}_n^H = 1 + \bar{R}_n^{SF} - 2 \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle} + O(\rho^n)$$

where $\rho < 1$ and $\bar{R}_n^{SF}$ is the average redundancy of the Shannon-Fano code, and as before

$$\alpha = \log_2 \left(\frac{1-p}{p} \right), \quad \beta = \log_2 \left(\frac{1}{1-p} \right).$$

Numerical Results for Huffman's Redundancy



(a) (b)

Figure 2: Huffman's code redundancy versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

Redundancy of Huffman's Code

Theorem 2. *Consider the Huffman block code of length n binomially(n,p) distributed over a binary alphabet. Then, for $p < \frac{1}{2}$ as $n \to \infty$*

$$\bar{R}_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\log 2} + o(1) \approx 0.057304 & \alpha \text{ irrational} \\ \\ \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n \rangle - \frac{1}{2}\right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle n\beta M \rangle / M} + O(\rho^n) & \alpha = \frac{N}{M} \end{cases}$$

where N, M are integers such that $\gcd(N, M) = 1$ and $\rho < 1$.

The **maximum Huffman redundancy** is

$$\max\{\bar{R}_n^H\} = 1 - \frac{1 + \log\log 2}{\log 2} = \lg(2(\lg e)/e) = 0.08607\ldots,$$

as $n \to \infty$.

Fourier Analysis and Sequence Distribution Modulo 1

We consider here redundancy of the **Shannon-Fano** block code and the **Huffman** block code for a memoryless source generating a block of length n with the binomial distribution.

Let $p(k) = p^k(1-p)^{n-k}$, where p is the probability of generating 0. Redundancy of the **Shannon-Fano** code is

$$\begin{aligned} \bar{R}_n^{SF} &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left(\lceil -\lg p(k) \rceil + \lg p(k) \right) \\ &= 1 - \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle \end{aligned}$$

$\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x and

$$\begin{aligned} \alpha &= \lg\left(\frac{1-p}{p}\right), \\ \beta &= \lg\left(\frac{1}{1-p}\right). \end{aligned}$$

Shannon-Fano Code – Irrational Case

Throughout, we shall use the following Fourier series; for real x

$$\begin{aligned}\langle x\rangle &= \frac{1}{2}-\sum_{m=1}^{\infty}\frac{\sin 2\pi m x}{m\pi}\\ &= \frac{1}{2}-\sum_{m\in\mathbf{Z}-\{0\}} c_m e^{2\pi i m x}, \qquad c_m=-\frac{i}{2\pi m},\end{aligned}$$

where $\mathbf{Z}$ is the set of all integers. Hereafter, we shall write $\sum_{m\neq 0} := \sum_{m\in\mathbf{Z}-\{0\}}$.

Irrational Case:

$$\begin{aligned}\bar{R}_n^{SF} &= \frac{1}{2}+\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\sum_{m\neq 0} c_m e^{2\pi i m(\alpha k+\beta n)}\\ &= \frac{1}{2}+\sum_{m\neq 0} c_m e^{2\pi i m\beta n}\left(p e^{2\pi i m\alpha}+q\right)^n.\end{aligned}$$

How to prove that the last term is $o(1)$ for α irrational?

Bernoulli Uniformly Distributed Sequences Modulo 1

Definition 1. [B-u.d. mod 1] *A sequence $x_n \in \mathbf{R}$ is said to be Bernoulli uniformly distributed modulo 1 (in short: B-u.d. mod 1) if for $0 < p < 1$*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \chi_I(\langle x_k \rangle) = \lambda(I)$$

holds for every interval $I \subset \mathbf{R}$, where $\chi_I(x_n)$ is the characteristic function of I (i.e., it equals to 1 if $x_n \in I$ and 0 otherwise) and $\lambda(I)$ is the Lebesgue measure of I.

Theorem 3. *Let $0 < p < 1$ be a fixed real number and suppose that the sequence x_n is B-uniformly distributed modulo 1. Then for every Riemann integrable function $f : [0,1] \to \mathbf{R}$ we have*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle) = \int_0^1 f(t)\, dt,$$

where the convergence is uniform for all shifts $y \in \mathbf{R}$.

Proof. Standard; cf. Drmota and Tichy (1997) or Kuipers and Niederreiter (1974).

Weyl's Criterion

Theorem 4. [Weyl's Criterion] *A sequence x_n is B-u.d. mod 1 if and only if*

$$\lim_{n\to\infty}\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}e^{2\pi i m x_k}=0$$

holds for all non-zero $m \in \mathbf{Z}-\{0\}$.

Proof. The proof again is standard. Basically, it is based on the fact that by Weierstrass's *approximation theorem* every Riemann integrable function f of period 1 can be uniformly approximated by a trigonometric polynomial (i.e., a finite combination of functions of the type $e^{2\pi i m x}$).

Finishing the Irrational Case

In our case, we must show that $\langle \alpha k \rangle$ is B-u.d. mod 1. By Weyl's criterion

$$\begin{aligned}\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{2\pi i m(k\alpha)} &= \lim_{n\to 0} \left(pe^{2\pi i m\alpha} + q\right)^n \\ &= 0\end{aligned}$$

provided α is irrational. Hence, by the previous theorem, with $f(t) = t$ and $y = \beta n$, we immediately obtain

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle = \int_0^1 t dt = \frac{1}{2}.$$

This proves that for α irrational

$$R_n^{SF} = \frac{1}{2} + o(1).$$

It can be proved (thanks to M. Drmota) that for almost all irrational α the rate of convergence in the above is $O\left(\frac{\log^{1+\delta} n}{\sqrt{n}}\right)$ for some $\delta > 0$.

Shannon-Fano Redundancy – Rational Case

Now we assume that $\alpha = N/M$ where N, M are integers such that $\gcd(N, M) = 1$. Denote $p_{n,k} = \binom{n}{k} p^k q^{n-k}$. We proceed as follows:

$$
\begin{aligned}
S_n &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left\langle k \frac{N}{M} + \beta n \right\rangle \\
&= \sum_{\ell=0}^{M-1} \sum_{m:\; k=\ell+mM \le n} p_{n,k} \left\langle \ell \frac{N}{M} + N + \beta n \right\rangle \\
&= \sum_{\ell=0}^{M-1} \left\langle \frac{\ell}{M} + \beta n \right\rangle \sum_{m:\; k=\ell+mM \le n} p_{n,k}.
\end{aligned}
$$

We will prove that the last sum is well approximated by $1/M$.

Useful Lemma

Lemma 1. *For fixed $\ell \leq M$ and M, there exist $\rho < 1$ such that*

$$\sum_{m:\ k=\ell+mM\leq n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{M} + O(\rho^n).$$

Proof. Let $\omega_k = e^{2\pi i k/M}$ for $k = 0, 1, \ldots, M-1$ be the Mth root of unity. It is well known that

$$\frac{1}{M}\sum_{k=0}^{M-1} \omega_k^n = \begin{cases} 1 & \text{if } M|n \\ 0 & \text{otherwise.} \end{cases}$$

where $M|n$ means that M divides n. In view of this, we can write

$$\begin{aligned} \sum_{m:\ k=\ell+mM\leq n} \binom{n}{k} p^k q^{n-k} &= \frac{1 + (p\omega_1 + q)^{n-\ell} + .. + (p\omega_{M-1} + q)^{n-\ell}}{M} \\ &= \frac{1}{M} + O(\rho^n), \end{aligned}$$

since $|(p\omega_r + q)| = p^2 + q^2 + 2pq\cos(2\pi r/M) < 1$ for $r \neq 0$.

Finishing the Rational Case

Continuing the derivation and using the above lemma we obtain

$$
\begin{aligned}
S_n &= \frac{1}{M}\sum_{\ell=0}^{M-1}\left(\frac{1}{2}-\sum_{m\neq 0} c_m e^{2\pi i m(\ell/M+\beta n)}\right) \\
&= \frac{1}{2}-\sum_{m\neq 0} c_m e^{2\pi i m n\beta}\frac{1}{M}\sum_{\ell=0}^{M-1} e^{2\pi i m\frac{\ell}{M}} \\
&= \frac{1}{2}-\frac{1}{M}\sum_{m=kM\neq 0} c_{kM} e^{2\pi i k M\beta n} \\
&= \frac{1}{2}-\frac{1}{M}\left(\frac{1}{2}-\langle \beta n M\rangle\right).
\end{aligned}
$$

Huffman Redundancy

We only need to analyze

$$T_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle}.$$

The **irrational** case is easy since a direct application of our previous result, with $f(t) = 2^{-t}$ and $y = \beta n$, yields

$$\lim_{n\to\infty} T_n = \int_0^1 2^{-t} dt = \frac{1}{2 \log 2}$$

for α irrational.

Huffman Redundancy – Rational Case

We could use Fourier series again, but instead we generalize our previous approach in the following theorem (proposed by M. Drmota):

Theorem 5. *Let* $0 < p < 1$ *be a fixed real number and suppose that* $\alpha = \frac{N}{M}$ *is a rational number with* $\gcd(N, M) = 1$. *Then, for every bounded function* $f : [0, 1] \to \mathbf{R}$ *we have*

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle k\alpha + y \rangle) = \frac{1}{M} \sum_{l=0}^{M-1} f\left(\frac{l}{M} + \frac{\langle My \rangle}{M}\right) + O(\rho^n)$$

uniformly for all $y \in \mathbf{R}$ *and some* $\rho < 1$.

Setting $f(t) = 2^{-t}$ we obtain

$$\begin{aligned} T_n &= \frac{1}{M} \sum_{l=0}^{M-1} 2^{-(l/M) - (\langle M\beta n \rangle / M)} + O(\rho^n) \\ &= \frac{1}{M} 2^{-\langle M\beta n \rangle / M} \frac{1 - 2^{-M/M}}{1 - 2^{-1/M}} + O(\rho^n) \\ &= \frac{1}{2M(1 - 2^{-1/M})} 2^{-\langle M\beta n \rangle / M} + O(\rho^n) \end{aligned}$$

for $\rho < 1$.

BONUS: Golomb's Code Redundancy

We consider **Golomb's code** that can be viewed as a special case of Huffman's code adapted to infinite alphabets distributed as $Geomteric(\theta)$, i.e.,

$$P(i) = (1-\theta)\theta^i, \quad i \in \mathbf{Z}^+,$$

is the probability assignment on the set of nonnegative integers where $0 < \theta < 1$.
Golomb proposed the following optimal binary code. Let ℓ be an integer such that

$$\theta^\ell = \frac{1}{2}.$$

In Golomb's code an integer i is represented as $i = \ell j + r$, where $j = \lfloor i/\ell \rfloor$. We encode j by a unary code (i.e., j zeros followed by a one) while r is encoded by a Huffman code (for a uniform distribution on ℓ symbols).

Golomb's Code Length and Redundancy

The code length $\mathbf{E}[L]$ is

$$\mathbf{E}[L] = \lceil \log_2 \ell \rceil + \frac{\theta^{2^{\lfloor \log_2 \ell \rfloor + 1}}}{\theta^{\ell} - 1},$$

The average redundancy $\bar{R}_{\ell}^{G}$ is

$$\bar{R}_{\ell}^{G} = \mathbf{E}[L] - H(\theta)$$

where the entropy $H(\theta)$ is

$$H(\theta) = -\log_2(1-\theta) - \frac{\theta}{1-\theta}\log_2 \theta.$$

Theorem 6. *Consider the Golomb code over an infinite alphabet generated by a source $Geometric(\theta)$. Then as $\theta \to 1$ or equivalently $\ell \to \infty$*

$$\bar{R}_{\ell}^{G} = 1 - \langle \log_2 \ell \rangle + 4 \cdot 2^{-2^{1-\langle \log_2 \ell \rangle}} - \log_2(\log_2 e) - \log_2 e - \frac{1}{2\ell} + O(\ell^{-2})$$

where, as before, $\langle x \rangle$ denotes the fractional part of x. Furthermore, the average redundancy $\bar{R}_{\ell}^{G}$ oscillates around $2 - \log_2(\log_2 e) - \log_2 e = 0.028538562\ldots$ with

$$\begin{aligned} \min\{\bar{R}_{\ell}^{G}\} &= 0.0251005712\ldots, \\ \max\{\bar{R}_{\ell}^{G}\} &= 0.0327344112\ldots. \end{aligned}$$

Sketch of Golomb's Redundancy

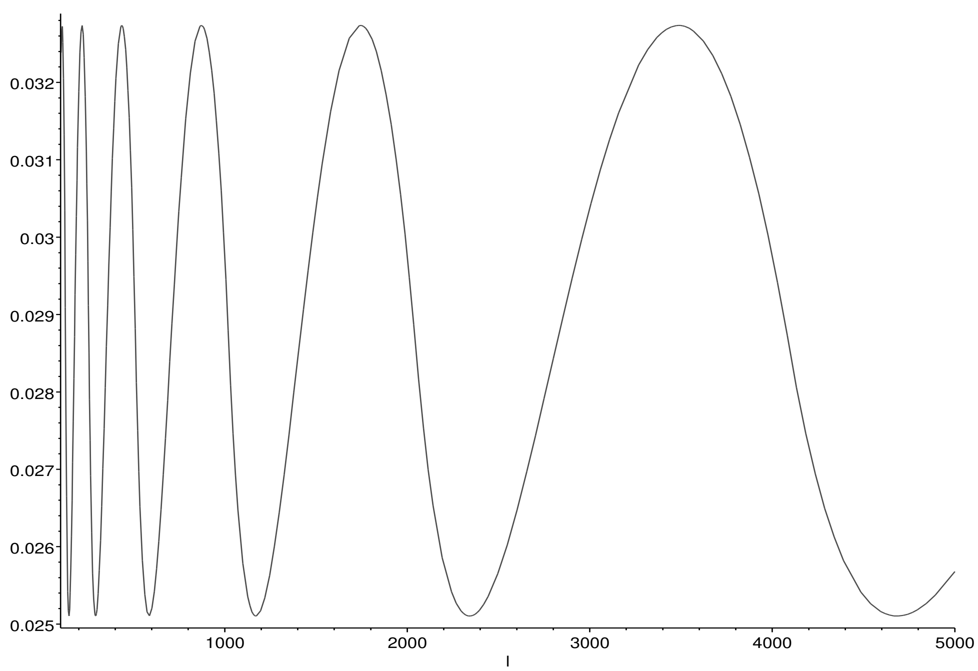


Figure 3: The average redundancy of Golomb codes versus ℓ.